The 2020 Nomination as Product of 2016’s Narratives

Corporations versus Climate Change: Examining the Implementation of Various Political Strategies by the Fossil Fuel Industry from 2008-2018

Dangerous Deception: Using Litigation to Minimize the Medical Misinformation Disseminated by Crisis Pregnancy Centers

Chinese Urbanism as a Way of Life: Becoming Urban in Shenzhen’s Informal Settlement

The Effect of Power Asymmetry on Regional Cooperation: Comparing Environmental Negotiations in ASEAN, the EU, and Mercosur

The Future of Economic Geopolitics: Network Effects in Intercultural Trade

Tunisia’s Ennahda Party: Developing a Framework for Interpreting Political Decision-Making in Historical Context

Photographs as Influencers: Searching for a Connection Between Visuals and Support for Legal Rights for Nature

Exploring the Rise and Fall of Bitcoin Prices: To What Extent was Optimistic Herding Behavior in Bitcoin Markets Disrupted by the Introduction of Bitcoin Futures Contracts?
### CONTENTS

**INTRODUCTION**

Editor’s Note  
Ciara O’Muircheartaigh  

**GUEST ESSAY**  
The 2020 Nomination as Product of 2016’s Narratives  
Seth Masket  

**ESSAYS**

**Corporations versus Climate Change:**  
Examining the Implementation of Various Political Strategies by the Fossil Fuel Industry from 2008-2018  
Daria Wick  

Dangerous Deception: Using Litigation to Minimize the Medical Misinformation Disseminated by Crisis Pregnancy Centers  
Holly Scrugham  

Chinese Urbanism as a Way of Life: Becoming Urban in Shenzhen’s Informal Settlement  
Yunhan Wen  

The Effect of Power Asymmetry on Regional Cooperation: Comparing Environmental Negotiations in ASEAN, the EU, and Mercosur  
Ann-Kathrin Merz  

The Future of Economic Geopolitics: Network Effects in Intercultural Trade  
Joshua Benjamin Curtis  

Tunisia’s Ennahda Party: Developing a Framework for Interpreting Political Decision-Making in Historical Context  
Stephen Garrett  

Photographs as Influencers: Searching for a Connection Between Visuals and Support for Legal Rights for Nature  
Elza Bouhassira  

Exploring the Rise and Fall of Bitcoin Prices: To What Extent was Optimistic Herding Behavior in Bitcoin Markets Disrupted by the Introduction of Bitcoin Futures Contracts?  
Manas Rajasagi
EDITOR’S NOTE

THE VIEW FROM SOUTH LAWN

The publishing process for the spring 2020 edition of the *Journal of Politics and Society* looked quite different from how it typically would. In the middle of our putting the edition together for print, the Editorial Board and I found ourselves scattered across the country and the world in the midst of a global pandemic, forced to collaborate virtually to bring the *Journal* to fruition.

This transition was personally challenging for all of us. For many members of our Editorial Board, myself included, spring 2020 was our final undergraduate semester at Columbia. To be ripped from campus and unable to graduate with our class was a heartbreaking experience. Even more heartbreaking, of course, was the knowledge that many of our peers were suffering immeasurable pain and loss at the hands of COVID-19.

Beyond these personal consequences, the pandemic has served as a reminder of the importance of the kind of work we strive to produce in the *Journal*. Innovative, methodologically rigorous research is perhaps our strongest weapon against COVID-19 and future viral outbreaks. Our Editorial Board is honored to make our modest contribution to the academic research ecosystem, and we look forward to identifying and nurturing exceptional undergraduate work in the semesters to come.

In the midst of countless cancellations and postponements due to COVID-19, this semester’s guest essay delves into one of the most important events whose logistics have been called into question: the 2020 United States presidential election. We are honored to feature the work of Seth Masket, professor of political science and director of the Center on American Politics at the University of Denver. In his essay for the *Journal*, Masket highlights key findings from his forthcoming third book, *Learning from Loss: The Democrats 2016-2020*, which considers the complex ways in which Democratic failure in 2016 paved the way for Joe Biden’s 2020 Democratic presidential nomination. Masket discusses the competing narratives surrounding Hillary Clinton’s 2016 loss, examining electability as a key point of accord. Masket’s work provides fascinating insight into the strategy of party nominations, brilliantly setting the stage for what is likely to be one of the most consequential elections in American history.

The eight works of undergraduate research published in this edition of the *Journal* represent our perennial commitment to methodological rigor and diversity of discipline within the social sciences. Through both qualitative and quantitative analysis, our authors consider a wide array of domestic and international social, political, and economic questions:

On the domestic front, Daria Wick explores the role of the fossil fuel industry in widening the partisan gap on the issue of climate change in the United States. Holly Scrugham considers potential legal strategies available for the regulation of crisis pregnancy centers in the US, outlining several approaches that may have greater impact than the popular mandated disclosure ordinance angle.

Outside of the United States, Yunhan Wen considers Chinese urbanism as a social—rather than a legal or geographic—construct through an ethnographic study of Baishizhou, an informal settlement in Shenzhen. Stephen Garrett leverages Tunisian political history to better understand and contextualize the Ennahda party’s contemporary decision-making and political strategy.

Still other papers consider international questions. Through his analysis of trade among three dozen Middle Eastern and East Asian countries, Joshua Benjamin Curtis examines the role of network effects in intercultural trade. Ann-Kathrin Merz studies power asymmetry in regional cooperation through an analysis of environmental negotiations in the Association of Southeast Asian Nations, Mercosur, and the European Union. Elza Bouhassira assesses photographic imagery as a tool to help secure legal personhood for nonhuman animals. Finally, Manas Rajasagi discusses Bitcoin’s collapse through the lens of behavioral economics.

Of course, we would not be able to share any of these exemplary works with you without support from the small army of people who help make the *Journal* possible each semester. On behalf of the entire Editorial Board, I thank our academic and administrative advisors, as well as our financial benefactors, for their continued support. Their time, encouragement, wisdom, and myriad other contributions enable our publication to grow and contribute to the Columbia
community in increasingly meaningful ways.

In closing, I would like to say a special thank you to the many exceptional people I have worked with throughout my three-and-a-half years on the Journal’s Editorial Board. I had no idea when I joined as an Associate Editor at the beginning of my first semester at Columbia that I was entering a community that would become a defining part of my undergraduate experience. You have all taught me so much, and I am so hopeful for the future knowing that you will be making waves in it. Onward.

Ciara O’Muircheartaigh

Editor in Chief

New York City

May 2020
The 2020 Nomination as Product of 2016’s Narratives

Seth Masket

The 2020 Democratic presidential nomination contest functionally ended on April 8th with the withdrawal of Sen. Bernie Sanders, leaving former Vice President Joe Biden the de facto nominee. It was a relatively quiet ending for the contest amidst a backdrop of pandemic and economic catastrophe, but it was fairly a remarkable outcome. Biden – a relatively moderate 77-year old white man with half a century of political experience – was somehow the last candidate standing in a contest involving the largest, most diverse, and arguably most liberal field of candidates in the Democratic Party’s history. Understanding how the nomination of Joe Biden came about requires an understanding of the election that preceded it and the way Democratic Party insiders interpreted it.

In my forthcoming book *Learning from Loss: The Democrats 2016-2020*, I follow discussions among influential Democrats – activists, Democratic National Committee members, officeholders, and others – from shortly after the 2016 election until the first state contests in early 2020. I find three main features of these discussions:

1. Conversations about whom to nominate for president in 2020 were heavily influenced by narratives about why the party’s presidential ticket lost in 2016.
2. There were a great many conflicting narratives that emerged to explain the 2016 election outcome, but a number of them suggested that Democrats would have a better chance of succeeding in 2020 by nominating a white, moderate man for the presidency.
3. Joe Biden’s nomination was hardly a fluke or a stroke of luck for him – the bulk of the party had rallied behind him long before the Iowa Caucuses, and much of the rest followed suit after the South Carolina primary.
4. Unlike the Republican Party of 2016, the Democratic Party of 2020 demonstrated that it was capable of making a collective choice and seeing that choice through the early primaries and caucuses.

The Importance of Narrative

The book *The Party Decides* describes the “Invisible Primary” stage of a nomination cycle as something of a “conversation” among party insiders. They are trying to communicate with each other to determine two vital things for picking a nominee: what they actually want to achieve as a party, and who they think they can get elected to achieve it. That second discussion was an especially fraught one for Democrats in the wake of 2016, as it was an unusually disorienting election for political observers. Pundits, activists, and others broadly expected Hillary Clinton to defeat Donald Trump that year, and drew on extensive polling and history in reaching that conclusion. Trump’s narrow Electoral College win caused many observers to question previous assumptions about what electability meant and undermined their faith in polling and their own political judgment.

In trying to figure out who could win in 2020, Democrats drew extensively on post-election narra-
About 2016. These narratives are the stories we tell ourselves about why complex events happen and what they mean. We hear these sorts of arguments all the time – “She should’ve gone to Wisconsin,” “Bernie would’ve won,” etc., were common examples emerging in November 2016. Some landslide losses have produced narratives that the party’s previous nominee was too ideologically extreme, as happened after Walter Mondale’s 1984 loss to Ronald Reagan. Politicians may claim a “mandate” if they or their party does decisively and unexpectedly well in an election.2

An important feature of these narratives is that they often aren’t based on very much hard data. Lippmann asked if, when we vote, “Have we expressed our thoughts on the public policy of the United States? …Surely the cross on a piece of paper does not express them.” But politicians and parties crave these narratives, especially when they lose. Winning, notes Hershey, has a “fairly blunt, conservatizing effect on campaigners.” As long as they’re winning, they’re going to take the view that whatever they did prior to that win was at least partially responsible for it, and they’ll continue to do it. Conversely, those who lose an election will be very open to making changes the next time they run, figuring that at least one of the things they did last time around was responsible for their loss.

I sought to examine how these narratives developed among Democratic Party activists in the wake of the 2016 election. I spoke with 65 Democratic activists – including party staff and leaders, campaign workers, donors, and others – in Iowa, New Hampshire, Nevada, South Carolina, and Washington, DC. Initial interviews began in 2017, and I followed up repeatedly with both subsequent interviews and on-line surveys until early 2020, just prior to the Iowa Caucuses.

Over the course of these many conversations, the activists explained to me their understanding of why the 2016 presidential election ended as it did. I have grouped these responses into eight different categories, summarized here:

1. Campaign activity – Tactics deployed by the Clinton campaign were inadequate and inappropriate.
2. Campaign message – The messaging used by the campaign failed to resonate with the voters in the way intended.
3. Identity politics – The campaign reached out to women, racial minorities, and other underrepresented groups but failed to speak to working class whites.
4. Candidate traits – Hillary Clinton had personality or campaigning traits that were inadequate to the task, or Donald Trump had traits that were under-appreciated prior to the election.
5. Racism/Sexism – Voters and/or the media held to racist or sexist biases that undermined Hillary Clinton’s potential vote share.
6. Sanders/Stein – Other candidates’ activities, particularly those of Sen. Bernie Sanders and Jill Stein, drew support away from Hillary Clinton.
7. Exogenous events – Events outside the campaign’s control, particularly intervention by Russia or activities by FBI Director James Comey prior to the election, hurt Clinton’s election chances.
8. Mood of the electorate – 2016 was somehow an unusual year, in which voters were craving atypical outcomes.

I provide numerous examples of these arguments in the book. I will summarize the key lessons here, however. The first is that there was substantial disagreement over just why Hillary Clinton lost to Donald Trump. Activists offered multiple explanations (most suggested two or three), with little consensus on what the problem was. Narratives varied considerably based on the gender identity of the activist, with men tending to find fault in Clinton as a candidate and in the choices of her campaign, and women tending to largely exonerate Clinton but blame systemic racism and sexism and outside factors like Russian interference. Some Clinton supporters blamed a lack of support from Bernie Sanders, while Sanders supporters were more critical of Clinton.

A second important lesson was that activists remained confused about the reasons for Clinton’s loss as time went on. Unlike what Hershey found in her analysis of post-1984 narratives, there was even less consensus among Democratic activists about the causes of Clinton’s loss in my follow-up survey in 2019, as the figure below demonstrates. Exogenous factors, particularly Russia and persistent sexism, were more likely to be blamed in the latter conversations. A separate analysis of media coverage finds that news coverage, which tends to converge on a few common post-election narratives following presidential contests, remained divided in the year following the 2016 election.
A third lesson revolved around just what these narratives suggested the Democratic Party should do for 2020. Campaigns can be improved and different candidates can be changed without the party needing to dramatically rethink what it stands for. But several narratives – particularly the “identity politics” narrative, favored by roughly a third of activists – had a challenging set of consequences. Accepting that narrative required believing that the party had lost precisely because of its emphasis on promoting diversity and enhancing the power of underrepresented groups. A survey experiment conducted in the book suggested that Democrats exposed to this argument were more likely to want the party to nominate a moderate white man for the next presidential race, and this effect was mainly found among African Americans and white women.

A final ingredient in the Democrats’ nomination decision was negative partisanship. Out-party approval of the president has been trending steadily downward over the decades; 23 percent of Democrats approved of George W. Bush’s performance, 14 percent of Republicans approved of Obama’s, and only eight percent of Democrats approve of Trump’s. And Trump’s confrontational and norm-shattering presidential style, of course, has done much to court this animosity.

The decision over whom to nominate for president tends to be heavily informed by what the party believes voters will support based on the most recent election results. But persistent disagreement over that complicated the decision-making process. Partisans could largely only agree that they wanted Trump out of the White House and were willing to pay a high price to achieve that, and many who were normally embracing of the party’s goal of diversity saw scaling back on that goal as a way to achieve a victory.

Seen in this way, Biden was a logical choice for such a party. Surveys among Democratic voters and activists repeatedly showed that, even when they didn’t see Biden as their top candidate, they saw him as the most electable, and they were prioritizing electability to a far greater degree than they had in modern history. Biden was also, in some ways, a relatively easy choice for party insiders – he was popular among the party’s voters, performed well in general election matchup polls, was closely tied to the beloved Obama administration, and was heavily supported among the African American community. But he also represented something of a retreat for a party that had tried something new in 2016 and, in the eyes of many, had been punished for innovation.

Endnotes

Harcourt, Brace, pp. 56-7.
COPORATIONS VERSUS CLIMATE CHANGE: 
EXAMINING THE IMPLEMENTATION OF VARIOUS POLITICAL STRATEGIES BY THE FOSSIL FUEL INDUSTRY FROM 2008-2018

DARIA WICK, UNIVERSITY OF CHICAGO CLASS OF 2019

ABSTRACT
A little more than ten years ago, climate change was a bipartisan issue that drew Nancy Pelosi and Newt Gingrich together to the steps of the United States Capitol with joint promises to address the looming crisis. Now, as 500-year storms begin to occur annually, one side of the political aisle has had presidential campaigns focused entirely on climate change while the other side denies its existence altogether. The shift from bipartisanship to polarization on this political issue did not occur by accident—the fossil fuel industry took action to alter the political climate surrounding anthropogenic climate change’s validity. What is less understood is how fossil fuel companies accomplished this. This paper seeks to map out which political strategies the American fossil fuel industry employed in attempting to influence U.S. climate change-related legislation in their favor. My results show that fossil fuel companies have employed a variety of political strategies that shift over time, indicating that the application of these strategies is not a random occurrence but rather part of a strategic effort on the part of the fossil fuel industry to alter their political landscape.

In the past half-century, the United States government has largely rolled back its laissez-faire approach to businesses and has instead passed countless laws regulating businesses’ practices. Companies view these regulations as a challenge to the maximization of profits. In reaction to the government’s increased presence in the marketplace, companies have become increasingly present in the political realm, functioning as political actors who wield a disproportionate amount of political influence in comparison to American citizens. Companies’ political activities allow them to shape their own regulatory and political environment, many times at the expense of the broader public. Given the impact corporations have on American political life it is important that political scientists understand the practical role private companies play in American democracy; this paper seeks to uncover specifics of how corporations work to influence their political environments.

In July 1977, ExxonMobil commissioned research to determine the effect fossil fuel combustion would have on Earth. Their study determined that continued use of carbon-based fuels would increase global average temperatures, resulting in environmental disasters. Knowing that widespread knowledge of this information would likely cause their product sales to plummet, the fossil fuel industry began a political campaign to prevent the issue of climate change from gaining political salience. Since then, the fossil fuel industry has made efforts to warp their political environment to influence legislation in their favor. Despite the potentially cataclysmic effects of climate change on the global population, the United States government has passed no significant climate change-specific legislation. Moreover, a recent study found that nearly a quarter of Americans are either unsure about, or do not believe in, man-made climate change.

Climate change began to become a salient issue (i.e. important to voters) in American political discourse in the 1980s. It was during that decade that global warming made the front page of The New York Times and climate change was an issue in a presidential election. Though the number of Republicans who believed climate change threats were being exagge-
ated by the media had risen over time, as recently as 2008 Republican and Democratic representatives were making joint statements pledging to find a bipartisan way of fighting climate change. \(^{10}\) During the Obama administration, largely fueled by the Tea Party movement, the climate change issue transitioned from being quasi-bipartisan to being a completely partisan issue, on par with abortion or gun control. \(^{11}\) Now, as the effects of climate change are beginning to be experienced (largely as they were predicted by scientists), the Trump administration’s official stance remains that it is a non-issue, with many in the administration openly denying climate change as a problem at all. \(^{12}\)

As the political environment surrounding the climate change issue has evolved from a bipartisan issue to an intensely partisan one, it is almost undeniable that the actions of corporations have helped shape this evolution. These companies employ strategies to both match and alter their political environment. The strategies that the fossil fuel industry have employed provide an empirical way to explore how private companies chose to interact with their political environment. Each company within the fossil fuel industry has a unique set of political interests. However, given that all fossil fuel companies will cease to exist (at least in their current state) if major climate change mitigation efforts are taken, all fossil fuel companies share similar political pressures from climate change action. Corporations are strategic political actors; the strategies they choose to employ to affect their political environment are chosen because the corporation believes they will be the most effective method of enacting change. \(^{13}\)

There is, of course, the potential argument that the political actions taken by fossil fuel companies are not done to influence climate change politics, but to influence other areas of legislation that affect the fossil fuel industry. After all, climate change is not the only regulatory issue that affects fossil fuel companies. This paper seeks to test this argument—are the political actions of fossil fuel companies regarding climate change a deliberate attempt on the part of the fossil fuel industry to influence climate change politics, or do such political strategies occur by coincidence as a part of different political actions taken by the oil, coal and natural gas corporations?

To answer this question, I will look at the various political strategies employed by fossil fuel companies between the years 2008 and 2018. I hypothesize that over this period, changes in the implementation of strategies by fossil fuel companies occurred as the political environment surrounding climate change evolved. If these changes occurred in a manner that is statistically significant, the data collected will provide empirical evidence that the political strategies being employed by fossil fuel companies in areas pertaining to climate change are not accidents, but deliberate attempts to react to and manipulate the political environment surrounding climate change.

**LITERATURE REVIEW**

**CORPORATIONS AS POLITICAL ACTORS**

Businesses do not devote resources to playing politics for fun; they do so to alter their economic landscape in a way that benefits themselves. \(^{14}\) Because government regulation can dramatically impact a company’s ability to maximize their profits, the United States’ current economic landscape incentivizes corporations to become politically involved. Companies in regulated industries, such as the fossil fuel industry, who invest in politics see an increase in profits. \(^{15}\) Therefore, companies in regulated industries need to become politically engaged if they want to maximize their profits, which requires the development of political strategies and tactics. \(^{16}\)

In the most basic terms, political actors are entities who have some degree of political power and/or authority within a particular society and partake in activities that have the potential to influence political events. \(^{17}\) Since the 1970s, companies have become more involved in the legislative process, intentionally trying to influence public policies, regulations, and laws in an attempt to shape their regulatory environment and advance their own private goals. \(^{18}\) Corporations’ role in U.S. life as job creators gives them the ability to spend a significant amount of money on shaping their political world and to influence the views of the broader job-seeking public. \(^{19}\) As political actors, corporations have largely been successful at accomplishing their goals. \(^{20}\) Because of their status as economic elites (at least when compared to the median citizen), the political activities of corporations have a disproportionate influence over American politics. \(^{21}\) As key actors of advocacy politics, corporations heavily influence the shape and content of government regulations. \(^{22}\)

**CLIMATE CHANGE AS A POLITICAL ISSUE**

Human-caused climate change is a looming threat to the continued survivability of the earth. There is little
scientific doubt that fossil fuel emissions dramatically affect the Earth’s climate, causing (among other things) higher global average temperatures, an increase in severe levels and rising sea levels.\textsuperscript{23} If human fossil fuel emissions continue at their current rate, the world is set to reach $2^\circ C$ of warming by 2100.\textsuperscript{24} Such an increase in global temperature would almost certainly be disastrous. Extreme heat waves, famine, water shortages, disease outbreaks, massive species extinctions, and severe weather events would be just a sampling of the “new normal” for humans to contend with.\textsuperscript{25} Additionally, climate change is a salient issue among the American public; a 2017 poll found that 72 percent of American believe in anthropogenic climate change.\textsuperscript{26} Given the apocalyptic threat climate change poses, and the widespread acknowledgement of the issue, there is pressure on the government to do something to mitigate the effects of climate change.\textsuperscript{27} Such actions would undoubtedly threaten the livelihood, and possibly the existence, of fossil fuel companies whose profits are currently dependent on the continued use of carbon-based fuels.\textsuperscript{28} As such, it is in the best interest of fossil fuel companies to influence politics to prevent the government from taking necessary steps to mitigate climate change.\textsuperscript{29}

**AMERICAN POLITICAL HISTORY OF CLIMATE CHANGE FROM 2008-2018**

Washington politics on global warming between the years 2008 and 2018 are largely stories of a growing partisan divide over the issue of climate change coupled with a progressive increase in conservative Republican power. In the 2008 presidential election both John McCain and then-Senator Obama acknowledged climate change as a legitimate threat and had plans to address it.\textsuperscript{30} Around that time polls found that 77 percent of Americans believed that the Earth was warming, with 47 percent believing that warming was caused by humans.\textsuperscript{31} In November 2008, Obama was elected into office with a democratic majority in both the House and the Senate. In 2009, Democrats in the House introduced the American Clean Energy and Security Act (ACES).\textsuperscript{32} ACES was an emissions trading plan and the first time either the House or Senate had passed a bill intended to curb greenhouse gasses and mitigate climate change.\textsuperscript{33} In response to ACES, Americans for Prosperity, a political advocacy group funded partly by the fossil fuel industry, ramped up their efforts to discredit climate change science.\textsuperscript{34} Americans for Prosperity led the charge towards making climate change a more partisan issue and helped defeat the passage of ACES in the Senate.\textsuperscript{35} By 2010, polls found that a mere 36 percent of Americans believed in man-made climate change.\textsuperscript{36} It was also around this time that Citizens United v. FEC was decided, opening the floodgates for corporate political spending in politics.\textsuperscript{37} The 2010 election saw Republicans gain control of the House due in part to incumbent Democrats from coal-country districts losing their seats to climate change-denying Republican challengers.\textsuperscript{38} Similarly, some of the more environmentally progressive Republicans saw their funding from conservative groups dry up or faced primary challengers due to their support for climate change initiatives.\textsuperscript{39} Republicans ended the House Select Committee on Energy Independence and Global Warming and quickly consolidated their party message that climate change was a hoax.\textsuperscript{40} At this time, public belief in climate change was at 34 percent, an all-time low.\textsuperscript{41}

In 2012, President Obama was re-elected with a Democratic Senate and Republican House, at which time public belief in climate change had risen from its 2010 low to 42 percent.\textsuperscript{42} President Obama began taking executive action to push his climate change agenda, as opposed to trying to pass legislation. In 2013, Obama began implementing his Climate Action Plan, a set of executive orders that focused on cutting carbon emissions, preparing infrastructure for the effects of climate change, and making the United States a leader in curbing climate change.\textsuperscript{43} In 2014, Republicans gained control of both the House and the Senate while President Obama unveiled his Clean Power Plan, an act which was taken to Court by several Republican attorneys general.\textsuperscript{44} President Obama issued a memorandum on coal leasing on federal lands, freezing any new applications to do so.\textsuperscript{45} In the 2016 election, Republicans gained control of the House, Senate and Presidency. Despite the fact that 53 percent of Americans believed in man-made climate change, with over 80 percent of registered voters believing the government should do something about the climate change issue, President Trump began immediately rescinding many of Obama’s climate change policies.\textsuperscript{46} By 2018, 57 percent of Americans believed in man-made climate change with 85 percent of registered voters believing the government should do something about the problem.\textsuperscript{47}

**POLITICAL STRATEGY FORMATION**

Corporations do not have to react to every political
event. However, once a corporation has decided to become politically active, there are a series of choices they must make on how to shape the content and implementation of their political strategy.48

The first decision a firm must make is whether they will take a transactional or relational approach to political action. In taking a transactional approach, companies react to political events on an issue-by-issue basis, typically waiting until an issue has become salient before developing a political strategy to address it.49 A transactional approach, in other words, is a reactive approach to political strategy; a company employing a transactional approach to forming political strategies will monitor public interest and only become involved in political events for specific issues.50 A transactional approach, thus, is better suited for companies who are not heavily impacted by government actions, who have a diverse set of products, or who operate in a more pluralist country.51 Conversely, companies can employ a relational approach to political strategy formation. Under a relational approach, companies develop a political network and employ political strategies broadly over the long term.52 In applying a relational approach, when a political issue becomes salient (or looks like it may become salient) corporations already have a political infrastructure in place to address the issue.53 This approach to political strategy formation also has the added benefit of reducing the transactional costs of political participation for the corporation.54 A relational approach to political strategy formation is better suited for companies, such as those in the fossil fuel industry, who are heavily impacted by government regulation, who do not have a diverse set of products, and who operate in a more corporatist country.55

Once a company has decided on their approach to political strategy they must then decide on their level of participation in the political arena.56 A company must decide if they want to act independently or work collectively with other companies.57 This decision is made by a combination of the company’s resources, the salience of the issue, whether the corporation is working in a corporatist or pluralist country, the type of industry in question, and the risk of collective action problems.58 Regardless of which approach to a political strategy is chosen, companies with greater resources are more likely to operate as individuals.59 However, more importantly than the resources of the firm or the structure of the government is whether or not the issue at hand is a salient one; companies are more likely to work collectively when dealing with a salient public issue.60 There are two likely reasons for this phenomenon. Firstly, working collectively minimizes each company’s exposure and liability if their actions provoke public outcry.61 Secondly, salient public issues are the kinds of issues that demand a large coalition to enact desired change, and individual companies typically lack the resources to amass such a coalition.62

Finally, once an approach to political strategy has been selected and the level of participation has been chosen, a firm is able to determine its specific political strategy (or strategies) and subsequent tactics. There are four political strategies companies can choose from: informational strategies, financial incentives, constituency building, or non-market strategies.63 Information strategies seek to target political decision makers by providing them with information. Financial incentive strategies aim to influence political decision makers by financially incentivizing them. Constituency building strategies affect political decision makers through their constituents. Non-market strategies aim to influence legislation after it has been passed.64

CONSTITUENCY BUILDING
Constituency building is a vitally important political strategy for corporations, consistently ranking as the most effective way a business can exert political influence while also providing corporations with a springboard for efficiently and effectively implementing other political strategies.65

Political decision makers’ primary goal is to be elected, a feat that requires people to vote for them. As a result, politicians are highly responsive to their constituents’ preferences.66 Constituency building works to align those constituent preferences with the policy goals of a corporation. Policy makers cannot act upon constituency interests of which they are unaware. Because more traditional methods of gauging constituent interests (polls, surveys, etc.) generally do not give the most accurate picture of what constituents’ interests are or how strongly they feel about those interests, policy makers tend to gauge constituent interests from constituent feedback.68 Moreover, because a small percentage of constituents are actually vocal, active participants in the political arena can have a dispro-
portionate impact on the decisions of policy makers.\textsuperscript{69} In layman’s terms, having a few constituents to “yell loudly” about an issue can make it seem to a policy maker that the issue at hand is important to voters, even if the larger body of constituents feels otherwise.

Constituency building is most effective for companies with straightforward political stances as opposed to more technical policy issues.\textsuperscript{70} In the case of climate change, for example, it may be easier to build a constituency who denies climate change exists as a problem, rather than build a constituency who has the technical knowledge to advocate for policy solutions to climate change. Additionally, the effectiveness of constituency building is impacted by the degree to which policy makers face electoral pressures from constituents.\textsuperscript{71} Practically, this means that policy makers who have larger bodies of constituents or who serve longer terms (such as Senators) feel less pressure from voters than policy makers with fewer constituents or shorter terms (such as House Representatives). Additionally, constituency building is not a one-size-fits-all strategy for corporations. The unique characteristics of a constituent body impact how a policy maker is influenced by corporate constituency building.\textsuperscript{72} Thus, because many corporations’ constituency building strategies require appealing to a broad constituent base, it is in a company’s best interest to have a versatile constituency building strategy that can work effectively across a large spectrum of voters.\textsuperscript{73}

Though companies can create artificial grassroots campaigns (more commonly known as “Astroturf campaigns”) to give the false appearance of popular appeal, constituency building is most effective when it stems from a place of genuine grassroots support.\textsuperscript{74} As such, an effective constituency building strategy involves efforts to control the information received by lay people.\textsuperscript{75} Political issues become salient through an “information cascade,” where issues deemed important by experts are picked up by the media, which are then picked up by the general population.\textsuperscript{76} It is in a corporation’s best interest, then, to skew the information cascade towards whichever viewpoint they want the public to take.

Strategies utilized by companies to affect the public’s information intake can be divided into two groups: one where companies avoid allowing an issue to become widely salient, and one where they react to an already widely salient issue.\textsuperscript{77} While an issue is still not widely known by the public, companies are best suited to try to “kill” an issue before it becomes more widely known by affecting how the issue is perceived by experts.\textsuperscript{78} An effective way to do this is to prevent experts from reaching a concurring opinion on the issue.\textsuperscript{79} Crucially, firms want to influence the information cascade at the expert level, while not influencing it so much that everyday people become aware of the issue.\textsuperscript{80} Conversely, for issues that are already widely salient, corporations should attempt to provide information to experts, reporters, or the public that support policies that are beneficial to the company.\textsuperscript{81}

There are numerous tactics a corporation can employ under the umbrella of constituency building, and a growing body of evidence shows fossil-fuel industries engaging in such tactics. Fossil fuel companies have poured money into attempts to affect the information cascade by funding research to foster uncertainty around man-made climate change.\textsuperscript{82} These industries have simultaneously funded Astroturf campaigns to give the impression of grassroots support for their cause and have utilized marketplace advocacy strategies with the intention of sowing uncertainty about climate change into the broader public while simultaneously increasing public support for fossil fuel companies.\textsuperscript{83}

**FINANCIAL INCENTIVES**

Corporations also utilize financial incentives to influence policy makers.\textsuperscript{84} Financial incentivization as a political strategy occurs when a corporation contributes financially to a policy maker. Because bribery remains illegal in the United States, a common financial incentive strategy takes the form of campaign contributions.\textsuperscript{85} Fossil fuel industries donate heavily to politicians. During the 2012 election cycle, for example, oil and gas companies spent over $70 million in political contributions.\textsuperscript{86} This seems to be money largely well spent; a 2017 study found that for every $10,000 a representative received from a climate change countermovement industry (of which fossil fuel companies are a part), the chances of that legislator voting in a pro-environmental manner decreased significantly.\textsuperscript{87}

For legislators in office, campaign fundraising is an ongoing event. Corporations can leverage this fact by contributing financially to candidates’ re-election campaigns in an attempt to influence their legislative behavior.\textsuperscript{88} It is important to note, however, that financial contributions from companies do not necessarily “buy” votes from legislators. Both lawmakers and businesses have incentives to ensure that the larger
public does not view corporate financial contributions as egregious bribery.\textsuperscript{89} Instead of corporate contributions existing as a quid pro quo for specific voting behavior, the relationship between corporate contributions and legislative activities is more of an “exchange of gifts.”\textsuperscript{90} Under this model, donations act as a favor from the corporation to the candidate, a favor which is expected to be returned if such a situation arises.

In the early stages of legislation (agenda setting, policy drafting, committee meetings, etc.), financial contributions exert the greatest amount of influence.\textsuperscript{91} Once a piece of legislation is up for a vote, the amount that legislative voting behavior is impacted by prior financial contributions is influenced by a myriad of factors. For example, corporate financial contributions are typically more effective for making regulatory changes and alterations to the tax code, but less effective for corporations seeking to affect other types of legislation.\textsuperscript{92} Additionally, regardless of how much money is thrown at lawmakers, legislation backed by strong public interest is less susceptible to being influenced by contributions than legislation that lacks public interest.\textsuperscript{93}

Corporations also contribute to non-incumbent political campaigns. Under this strategy, candidates can be viewed as something resembling investments for companies.\textsuperscript{94} The goal of contributing to a campaign is not necessarily to ensure that favorable candidates get elected, but rather to form a favorable political connection with a candidate in the event that they are elected.\textsuperscript{94} Politics is characterized by repeated, reciprocal favors; by donating to a campaign, corporations are doing potential future lawmakers a “favor,” one which firms expect to be reciprocated sometime in the future.\textsuperscript{96} While campaign funding does play a crucial role in whether or not a candidate will be elected, it is not in the candidate’s, nor in the company’s, interest to appear as though corporations are buying potential (or sitting) lawmakers. As with corporate financial contributions, campaign contributions are not a quid pro quo system. Rather, as with corporate financial contributions, they serve as a transaction in an “exchange of gifts” between firms and potential legislators.\textsuperscript{97} In other words, when contributing to campaigns, companies are not buying future politicians; they are buying political relationships that can be used in the future.

**INFORMATION STRATEGIES**

In a similar manner to their implementation of constituent building strategies, corporations will utilize information strategies to affect policy makers.\textsuperscript{98} While constituency building strategies aim to inform the public in a way that subsequently impacts the legislature, information strategies aim to directly inform policy makers in the hopes of impacting their decisions.

Perhaps the most well-known tactic under the information strategy umbrella is lobbying. Fundamentally, lobbying provides a policy maker with information to influence their political decision making on behalf of a third party.\textsuperscript{99} Lobbyists are paid to act as a “middle man” between whom they are lobbying for and government officials.\textsuperscript{100} Lobbying efforts attempt to influence the decisions of policy makers by impacting their understanding and perceptions of an issue.\textsuperscript{101} Lobbyists do this by providing overwhelming information for their preferred side of an issue, saturating policy makers with information and priming them to agree with arguments and information that are beneficial to the lobbyist.\textsuperscript{102} A crucial component in this information process is a social connection between the lobbyist and the policy maker.\textsuperscript{103} Information exchanges between lobbyists and policy makers form stable relationships based on a shared perception and common cultural orientation.\textsuperscript{104} These relationships create a stable network that lobbyists can use to maintain their influence and control the flow of information to policy makers.\textsuperscript{105} There is ample evidence of fossil fuel industries engaging in such efforts. From 2000 to 2016, such corporations spent almost $2 billion on climate change-related lobbying.\textsuperscript{106}

**NON-MARKET STRATEGIES**

After legislation has been passed, there are still many ways corporations can influence policy to their benefit.\textsuperscript{107} Non-market strategies target the judicial or executive branches of government to impact regulatory and legal environments in the hopes of mitigating the regulatory pressure that corporations act under.\textsuperscript{108} The goal of non-market strategies is to change the interpretation of existing legislation so that it no longer resembles, or no longer has the same effect, as it was originally intended.\textsuperscript{109}

Once legislation has been passed, it must be implemented, a job conducted by the executive branch of the United States government. Executive agencies are in charge of interpreting legislation and determining which regulatory policies to implement in order to enact a piece of legislation.\textsuperscript{110} In doing so, regulatory agencies are functionally selecting policy positions.
that can impact a corporation in various ways.\textsuperscript{111} If firms are able to influence the regulatory policies of an executive agency, they can impact the degree to which a piece of legislation positively or negatively affects their company.\textsuperscript{112} Because most regulators are appointed, they cannot accept financial contributions, nor do they feel constituent pressures like legislators do. Thus, in attempting to influence regulatory policies, most companies provide information to regulators in an attempt to influence them.\textsuperscript{113} Since the 2016 election, fossil fuel industries have been remarkably successful at altering their regulations at the agency level which has led to functionally fewer emissions regulations than could otherwise be implemented.\textsuperscript{114}

Alternatively, companies can use the courts to impact legislation. While the judicial branch works to maintain an image of impartiality, the fact of the matter is that judges are political actors who tend to render decisions, and thus alter policy, in accordance with their own respective political ideology.\textsuperscript{115} Companies can provide judges with information in the form of amicus curiae briefs in an attempt to lobby them.\textsuperscript{116} Attempting to use the judicial system to impact legislation, however, is a risky, often last-ditch, effort for corporations given the risk that such a strategy may backfire and set an unfavorable precedent.\textsuperscript{117}

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Characteristics</th>
<th>Tact</th>
<th>Related Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information</td>
<td>Influences political decision makers by providing information</td>
<td>- Lobbying</td>
<td>- Paying lobbying firm for climate change-related lobbying tactics directed at legislators</td>
</tr>
<tr>
<td>Financial Incentive</td>
<td>Targets political decision makers by providing financial incentives</td>
<td>- Financial contributions</td>
<td>- Paying individual lobbyist for climate change-related lobbying tactics directed at legislators</td>
</tr>
<tr>
<td>Constituency Building</td>
<td>Targets political decision makers through constituent support</td>
<td>- Advocacy advertising/Marketplace advocacy</td>
<td>- Official testimony by employees of fossil fuel company</td>
</tr>
<tr>
<td>Constituency Building</td>
<td>Astroturf campaigns</td>
<td>- Paying for legislators to attend conferences</td>
<td>- Funding research and using results to supply position papers or technical reports</td>
</tr>
<tr>
<td>Constituency Building</td>
<td>Informing employees</td>
<td>- Financial contributions to any politician who votes against climate change legislation</td>
<td>- Funding towards anything related to a legislator attending an informative conference on the political issue at hand</td>
</tr>
<tr>
<td>Constituency Building</td>
<td>Strategic litigation</td>
<td>- Financial contributions to any candidate who holds a publicly stated anti-climate change regulation stance</td>
<td></td>
</tr>
<tr>
<td>Nonmarket Strategies</td>
<td>Proactive outreach</td>
<td>- Advertising</td>
<td>- Press releases denying climate change's existence</td>
</tr>
<tr>
<td>Nonmarket Strategies</td>
<td>Public relations efforts</td>
<td>- Encouraging employees to mobilize as political actors in a specific way</td>
<td>- Press releases encouraging the public to form a specific opinion on climate change legislation</td>
</tr>
<tr>
<td>Nonmarket Strategies</td>
<td>Regulatory capture</td>
<td>- Contrarian scientists appearing in news media to cast doubt on the validity of climate change science</td>
<td></td>
</tr>
<tr>
<td>Nonmarket Strategies</td>
<td>Lobbying</td>
<td>- Corporate representatives appearing in news media to cast doubt on the validity of climate change science</td>
<td></td>
</tr>
<tr>
<td>Nonmarket Strategies</td>
<td>Amicus curiae briefs</td>
<td>- Providing funding for AstroTurf campaigns</td>
<td></td>
</tr>
</tbody>
</table>

\textbf{Table 1: Summary of Political Strategies}

\textbf{Table 2: Indicators of Political Strategies}

\textbf{DATA COLLECTION & ANALYSES}

For this paper, a “fossil fuel company” is defined as any company who produces or refines oil, coal or natural gas for energy use. The “fossil fuel industry”
is the aggregate of these companies. Analyses presented in this paper focus on parent companies active between 2008 and 2018 who have articles of incorporation within the United States. This time frame is examined because it provides an opportunity to study distinct political environments surrounding the climate change issue. From 2008 to 2012, the public’s opinions on climate change shifted significantly; the belief that humans were causing the climate to change decreased dramatically despite the legislature’s efforts to tackle climate change. From 2012 to 2016, however, public opinion on human-caused climate change began to rise again and was met with actions by the executive branch that worked to mitigate the effects of climate change. Finally, from 2016 to 2018, despite the fact that the majority of Americans believe the United States government should act to mitigate the effects of climate change, the Trump Administration has been actively pushing a pro-fossil fuel agenda. Thus, the time frame between 2008 and 2018 provides three distinct political environments that fossil fuel companies have operated in and provides an opportunity to examine how fossil fuel companies react as strategic political actors various political landscapes.

Because consumers do not like to see businesses meddling with their politics, an important characteristic of corporate political activity is the need to “hide” a company’s political actions. Therefore, it is not uncommon that companies work through trade organizations, PACs or think tanks when attempting to influence politics. To account for this, I consider trade group, think tank, and PAC membership or donorship as a continuation of a company’s political activities.

Table 1 outlines the potential political strategies fossil fuel companies could apply, as well as the subsequent tactics fossil fuel companies could employ to accomplish these strategies. Table 2 outlines the real-world indicators that these tactics are being used. It is important to note that each strategy has a different number of mutually exclusive tactics that could be employed to accomplish the strategic goal. Therefore, looking at the aggregation of tactics used for each strategy would provide a bad strategy-by-strategy comparison. For the purposes of data collection, the application of any one tactic under the larger strategy umbrella was considered as sufficient evidence that that strategy is being used. In an attempt to avoid skewing data towards political strategies with a larger number of possible tactics, data were collected by gathering information regarding whether or not a company employed a particular strategy at least once in a particular time period. By comparing the prevalence of strategies from year to year, I was able to examine how fossil fuel companies reacted to their changing political environment.

A list of fossil fuel companies was compiled by searching through SEC codes 1311 (crude petroleum and natural gas), 2911 (petroleum refining), 1220 (coal mining), 1221 (surface coal mining) and 5172 (petroleum products) for parent companies, which were then verified by conducting a Bloomberg and Google search of the company’s name. Companies who could not be verified as parent companies active for the duration of the time studied were not included on the list of fossil fuel companies. A list of think tanks and political advocacy groups that promote climate change denialism was compiled using reports from the Union of Concerned Scientists and Public Accountability Initiative. Their lists of corporate members or donors was compiled using end-of-year reports and IRS non-profit filings. A list of industry trade organizations were compiled using Google searches of “trade organizations + oil”, “trade organizations + coal” and “trade organizations + natural gas.” Trade organization member lists were then used to determine which companies within the pre-established list of companies were members of each trade organization.

To build this dataset, I utilized a variety of databases (mentioned below) as well as Google to find evidence of political tactics. When using databases, data was collected by searching companies by name and then, when relevant, using the terms “climate change”, “global warming”, “fossil fuels”, “greenhouse gas” and “greenhouse emissions” to further narrow down the search parameters. Anything tagged by Congress.gov as being related to “climate change” or “global warming” was considered to be climate change related legislation. Opensecrets.org was used to find evidence of companies lobbying regulators, lobbying legislators, making financial contributions or engaging in regulatory capture. Lexis.un was used to find evidence of companies participating in the regulator’s public notice. Instances of corporations, or their proxies, testifying as expert witnesses were collected by looking through House and Senate testimony records. Office of Congressional Ethics records were used to determine instances when companies, or
their proxies, paid for government officials to attend conferences on climate change issues. To determine if any companies had funded “climate change contrarian research” I Googled “climate change contrarian scientists,” made list of prominent climate change contrarian scientists, entered their names into Google scholar and JSTOR, found their papers, and then followed the money as far back as I could using the funding disclosures required by many journals as well as by checking with fossil fuel corporation’s SEC filings. To find evidence of corporate AstroTurf funding I searched “name’ + ‘AstroTurf campaign funding’” and “name’ + ‘AstroTurf’.” To find evidence of informing employees I used Google to search “name’ + ‘politically informing employees’” and “name’ + ‘politically influencing employees’.” To find possible instances of advocacy advertising I searched “name’ + ‘ad campaign’” in Google and YouTube. Finally, to find evidence of public relations efforts or impacting the information cascade, I searched “climate change” and “global warming” in company, or their proxies’, websites as well as Googled “name’ + ‘climate change denial.’” After all the data had been collected, the number of individual companies who employed a political strategy was aggregated within each of the segmented periods of time and for each of the four political strategies.

RESULTS
Of the 226 companies examined, 99 of them (43.8%) employed at least one type of political strategy at least once between November 4, 2008, and November 6, 2018. The prevalence of company political action varied from time period to time period: fluctuating dramatically twice, decreasing significantly between 2008 and 2010 and between 2010 and 2012 and then increasing dramatically between November, 2014 and November, 2016 as well as between November, 2016 and November, 2018. A chi-squared test of the fluctuations and rates of political strategies employed resulted in a p-value of 0.013, indicating that the choice of companies to implement political strategies is not random, but rather that companies are employing these strategies with the intention of affecting political change. Table 3 outlines the fluctuations in strategy use for five election cycles between 2008 and 2018. Table 4 outlines trends seen within each of the corporate political strategies.

Table 3

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Information Strategies</td>
<td>63</td>
<td>60</td>
<td>63</td>
<td>63</td>
<td>68</td>
<td>317</td>
</tr>
<tr>
<td>Financial Incentives</td>
<td>64</td>
<td>66</td>
<td>67</td>
<td>70</td>
<td>72</td>
<td>339</td>
</tr>
<tr>
<td>Constituency Building</td>
<td>74</td>
<td>64</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>327</td>
</tr>
<tr>
<td>Nonmarket Strategies</td>
<td>54</td>
<td>62</td>
<td>56</td>
<td>55</td>
<td>62</td>
<td>289</td>
</tr>
<tr>
<td>Total:</td>
<td>255</td>
<td>252</td>
<td>249</td>
<td>251</td>
<td>265</td>
<td></td>
</tr>
<tr>
<td>Total # of individual companies that employed at least one strategy</td>
<td>89</td>
<td>83</td>
<td>82</td>
<td>82</td>
<td>88</td>
<td>99</td>
</tr>
</tbody>
</table>

Table 4

<table>
<thead>
<tr>
<th>Year/Timeframe</th>
<th>Information Strategies</th>
<th>Financial Incentives</th>
<th>Constituency Building</th>
<th>Nonmarket Strategies</th>
<th>All Strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total # of companies who were politically active during every time frame</td>
<td>56</td>
<td>64</td>
<td>63</td>
<td>48</td>
<td>43</td>
</tr>
</tbody>
</table>

Table 5

<table>
<thead>
<tr>
<th>Year/Timeframe</th>
<th>Information Strategies</th>
<th>Financial Incentives</th>
<th>Constituency Building</th>
<th>Nonmarket Strategies</th>
<th>All Strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td>#companies who were politically active during only one timeframe</td>
<td>10</td>
<td>2</td>
<td>11</td>
<td>13</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 6

<table>
<thead>
<tr>
<th>Year/Timeframe</th>
<th>Information Strategies</th>
<th>Financial Incentives</th>
<th>Constituency Building</th>
<th>Nonmarket Strategies</th>
<th>All Strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td>#companies who employed a strategy at least once</td>
<td>76</td>
<td>72</td>
<td>75</td>
<td>72</td>
<td>99</td>
</tr>
</tbody>
</table>
DISCUSSION

GENERAL FINDINGS

The findings from this project demonstrate that the majority of fossil fuel companies do not actually engage in any political activity. This result, in itself, is not all that shocking. Though it is widely known that corporations affect American politics, studies have shown that the majority of companies are politically inactive.124 When it comes to corporate political activity, industries face the same collective action problems that plague many other groups.125 My goal was to examine whether the fossil fuel industry is being strategic in their application of political strategies in order to affect climate change policy. All fossil fuel companies benefit from efforts by any company to push climate change denialism because all fossil fuel companies would suffer from efforts to mitigate climate change. The fact that not all fossil fuel companies engage in political activity provides evidence suggesting that the fossil fuel industry is being strategic in their political actions - for some fossil fuel companies it is more strategic to sit back and let someone else expend resources promoting climate denial. The fact that the majority of fossil fuel companies are politically inactive does not mean that the industry, on the whole, is not actively trying to push climate change denialism. Rather, it indicates that funding climate change denial is a collective action problem for the fossil fuel industry. The shifts in the total number of politically active companies seen in Graph 2 lends further support to this—changes in the political environment correlate with shifts in the number of politically active companies, suggesting that as political actors, companies view certain political environments, but not others, as being worth the resources of becoming politically active themselves.

As seen in Graph 1, financial incentives were, with one exception, the most commonly employed strategy in every individual election cycle. Interestingly, financial incentive strategies were the only political strategy that had zero attrition, meaning that once a corporation began employing financial incentive strategies, they did not stop employing financial incentive strategies. As previously established in the literature review, financial incentive strategies are employed to build political relationships with legislators.126 The fact that financial incentive strategies were the most commonly employed political strategy between 2008 and 2018 suggests that relationships with politicians are an important part of the fossil fuel industry’s attempts to push climate change denial.

Nonmarket strategies were, with one exception, the least frequently used strategy within each time frame. This finding was not particularly surprising. Nonmarket strategies are attempts to affect legislation that has been passed and as such are used as
last-ditch efforts by companies. Because nonmarket strategies involve impacting laws that have already been passed (a position all other political strategies work to avoid) and they have a greater risk of backfiring with long-lasting ramifications, which makes them something of a Hail Mary political strategy for corporations. The finding that nonmarket strategies were, across the board, the least commonly used political strategy indicates two things. Firstly, it indicates that there has not been a whole lot of climate change legislation passed through the United States government. The fact that companies are employing nonmarket strategies significantly less than any other corporate political strategy shows that there has been little climate change legislation passed for fossil fuel companies to try to affect. Secondly, these results may suggest that the fossil fuel industry would prefer to stop legislation before it is passed, as opposed to trying to weaken it afterwards. This result also may suggest that preventing climate change legislation from passing is the objective of the fossil fuel industry.

POLITICAL CONTEXT FOR SPECIFIC FINDINGS
Statistical analysis of Table 3 indicates that the shifts in political strategies being employed by fossil fuel companies are not random, which suggests that fossil fuel companies have employed these political strategies deliberately to affect climate change-related legislation in their favor. This, of course, begs the question: what is the political context for the trends seen in the data?

In the 2008 presidential election, both then-Senator John McCain and then-Senator Barack Obama acknowledged climate change as a legitimate threat and had plans to address it. Around that time polls found that 77% of Americans believed that the earth was warming, with 47% believing that warming was caused by humans.

In early 2009, President Obama was sworn into office alongside a Democratic House and Senate. Later that year, Democrats in the House introduced the American Clean Energy and Security Act (ACES). The ACES was an emissions trading plan and the first time either the House or Senate had passed a bill intended to curb greenhouse gasses and mitigate climate change. In response to ACES, Americans for Prosperity, a political advocacy group funded partly by the fossil fuel industry, ramped up their efforts to discredit climate change science. Americans for Prosperity led the charge towards questioning climate science and making climate change a more partisan issue, effectively defeating the passage of ACES in the Senate. By 2010, polls had found that a mere 36% of Americans believed in man-made climate change. It is around the time of the ACES’ attempted passage through the legislature that the data shows an increase in the total number of politically active companies, as well as a peak in the number of companies who employed constituency building strategies.

Between 2008 and 2010 and between 2010 and 2012, the instances of constituency building strategies declined dramatically from 74 to 64 companies engaging in such strategies. When placed in the context of the battle over the ACES, this decline makes sense. The climate change denial movement had ramped up their efforts to make climate change a charged partisan issue in 2009, but when ACES was defeated the need for constituency building strategies was reduced. An alternative, but not mutually exclusive, explanation for the decrease in constituency building strategies is the public’s belief in climate change. In 2010, belief in climate change was at an all-time low. It is possible that companies stopped deploying constituency building strategies because they felt that such strategies had accomplished their goal and were no longer required. This explanation, however, does not account for why those companies did not reimplement such strategies once public acceptance of climate change began to grow again. If constituency building strategies were a reactionary effort of the fossil fuel industry to try to change public opinion, one would likely see an increase in the number of companies employing constituency building strategies as public opinion regarding the validity and threat of climate change grew. This assumption, however, is not what the data show. Rather, the results from my data collection suggest that constituency building strategies are used by the fossil fuel industry in response to legislative action, not in response to broader public opinion.

Another important political event that occurred between November 4th, 2008 and November 2nd, 2010 was the Supreme Court’s decision in Citizens United v. Federal Election Commission, 558 U.S. 310 (2009). The Court’s decision opened the floodgates to outside spending in politics. As a result, the 2010 election cycle saw fossil fuel interests dumping money into the election or re-election of politicians who were skeptical of climate change. Incumbent Democrats from coal-country districts lost their seats to climate change denying Republican challengers, while at the
same time, environmentally progressive Republicans saw their funding from conservative groups dry up or faced primary challengers. Given the observed data, it would appear as though the Citizens United decision did not affect how many companies employed financial incentive strategies. That being said, since the data collected does not provide an analysis of the use of financial incentive strategies for the election cycle prior to the Citizens United decision, it would be unwise to make that assumption. Citizens United did, however, come into being right before the number of companies employing financial incentive strategies dropped. It is plausible to think that, as Citizens United opened the door for dark money (political spending from nonprofit groups), fossil fuel companies saw financial incentive strategies as a better use of resources than constituency building. It would follow, then, that companies who had been employing both strategies stopped devoting those resources towards financial incentive strategies instead, thereby helping explain the drop in companies who employed constituency building strategies after November 2nd, 2010.

Between 2010 and 2012, Graph 1 shows a significant increase in the number of companies employing nonmarket strategies. This is likely due to the case Coalition for Responsible Regulation, Inc. v. EPA, 684 F.3d 102 (D.C. Cir. 2012). In June, 2012, the Coalition for Responsible Regulation, a group comprised of members of the fossil fuel industry, alongside a handful of state’s attorneys general brought the case Coalition for Responsible Regulation Inc. v. EPA to Federal Court. The aim of the lawsuit was to challenge the 2007 Supreme Court decision Massachusetts v. EPA, 549 U.S. 497 (2007), which had ruled that greenhouse gasses were air pollutants and could be regulated under the Clean Air Act. Since 2009, the Obama Administration’s EPA had been making moves to limit greenhouse gases to impede climate change. Because the EPA was using the Clean Air Act and Massachusetts v. EPA to justify its actions, nonmarket strategies provided the only possible route for the fossil fuel industry to prevent this instance of greenhouse gasses being regulated. Though the challenges brought up in Coalition for Responsible Regulation Inc. v. EPA were eventually dismissed by the federal appeals court, this lawsuit did involve numerous companies within the fossil fuel industry and was likely the reason the data shows a dramatic increase in the number of companies who employed nonmarket strategies between November 2008 and November 2010 and between November 2010 and November 2012.

In 2012, President Obama was reelected with a Democratic Senate and Republican House, at which time public belief in climate change had risen to 42% from its 2010 low. The President began taking executive action to push his climate change agenda, as opposed to using legislative methods. In 2013, Obama began implementing his Climate Action Plan, a set of executive orders that focused on cutting carbon emissions, preparing infrastructure for the effects of climate change and making America a leader in curbing climate change. Surprisingly, as the President was taking action to reduce greenhouse gases, the data analysis presented in this paper does not demonstrate significant increases in the number of companies employing any political strategies. There are a few explanations for this outcome. Perhaps the most simplistic explanation is that fossil fuel companies did not change their political strategies during this time period simply because doing so would have accomplished nothing. President Obama wasn’t taking legislative action, so employing financial incentives or information strategies to sway members of Congress would have accomplished nothing. Additionally, Obama was at his term limit so he was no longer constrained by reelection pressures; therefore, constituency building strategies would not have had any impact. This, of course, begs the following question: why did the number of companies employing nonmarket strategies not increase? As constituency building, information and financial incentive strategies would have been ineffective, nonmarket strategies should have obviously been the last-ditch effort on the part of the fossil fuel industry to stop Obama’s executive actions. Yet, between 2010 and 2012 and between 2012 and 2014 the number of companies employing nonmarket strategies decreased. It is possible that the reason nonmarket strategies declined during this time period is because, as with the other forms of political strategies, corporations simply thought employing them would be an ineffective use of resources. An alternative explanation is that nonmarket strategies were being employed to weaken Obama’s executive action, they were just not being taken by the fossil fuel industry as defined in this thesis. Between 2012 and 2014, the Utility Air Regulatory Group, an organization comprised primarily of electricity companies, as well as multiple states’ attorneys general, brought suit against Obama’s executive actions. Perhaps the reason that more fossil
fuel companies did not employ nonmarket strategies in response to the Executive’s Climate Action Plan was because other entities were already advocating against President Obama’s actions and so Fossil Fuel companies did not need to devote resources to such political actions.

In 2014, Republicans gained control of both the House and the Senate while President Obama unveiled his Clean Power Plan, an act which was taken to court by several Republican attorneys general. President Obama also issued a moratorium on coal leasing on federal lands, freezing any new applications. As with the time period between 2012 and 2014, there was very little change in the number of companies employing any of the political strategies between 2014 and 2016. The proportion of companies employing each respective type of political strategy remained holistically static. This is not all that surprising. The political climate surrounding the climate change issue remained largely unchanged between 2012 and 2014 and between 2014 and 2016, so it stands to reason that there would not be any significant changes in the number of companies who employed each form of political strategy.

In the 2016 election, Republicans gained control of the House, Senate and presidency. President Trump began immediately rescinding many of Obama’s climate change policies while pushing for an increased use in fossil fuel energy sources. The data shows that between 2014 and 2016 and between 2016 and 2018 there was a significant increase in the total number of companies who were politically active. The increase in the number of companies employing nonmarket strategies was likely a response to how fossil fuel friendly the executive administration became after Trump’s inauguration. The Trump Administration’s nominated a large number of people with connections to the fossil fuel industry to positions within the Environmental Protection Agency and Department of the Interior; all while implementing policies intended to increase the use of fossil fuels. It would seem as though more companies employed nonmarket strategies because, with a pro-carbon president, such strategies are likely to affect the kind of change the fossil fuel industry wants. Indeed, in this paper I have found that from 2014 to 2016 and from 2016 to 2018, there were noteworthy increases in the number of companies who employed information strategies. This is interesting because, during this time period, the legislature did not introduce, let alone pass, any major legislation that would have benefitted the fossil fuel industry. One would think that with an increase in information strategies and a pro-carbon party controlling both the House and the Senate, some legislation would have been passed to benefit the fossil fuel industry. A potential explanation for this phenomenon is that the use of information strategies increased in order to influence politicians against public opinion on climate change. In 2016, 53 percent of Americans believed in man-made climate change, with over 80 percent of registered voters believing the government should do something about the climate change issue. By 2018, 57 percent of American believed in man-made climate change with 85 percent of registered voters believing the government should do something about the problem. It is possible that, in the face of growing public opinion against the fossil fuel industry’s interests, companies began to employ information strategies to prevent legislatures from enacting the change their constituents support. This would suggest, of course, that the fossil fuel industry has recently stopped trying to outright convince the broader public that climate change is not happening, and they are more focused on making sure change happens as slowly as possible. However, given the fact that extreme weather events are making climate change harder to blatantly deny, thereby rendering constituency building strategies less effective, it is very plausible that the increase in information strategies was an attempt to counter growing public demand for action on climate change.

LIMITATIONS
This study had several limitations. Because of the time and resource limitations of this project I relied exclusively on public data collection sources for my data collection. While these sources varied from databases to search engines to websites, they were all ways of collecting information that are open to the broad public. Because of the backlash companies can face if they seem to be too heavily engaged in politics, there...
is a significant incentive for companies to hide their political activities. It is likely, then, that at least some companies are politically active in ways that cannot be found using public data collection sources. Therefore, those activities would not have been accounted for in my data. Moreover, my results relied heavily on the use of Google. Initially, as I was gathering data, I ran keywords through multiple search engines (Ecosia, Google, Bing and Yahoo). However, I quickly realized that Google provided the same results as the other search engines were, and so, for the efficiency’s sake, I switched to only using Google. Additionally, this work only looked at the actions taken by the company, ignoring political actions taken by individual high-ranking members of companies, and assuming that the company conducted its political actions independently of the political activity of high-ranking members of that company. For publicly traded companies, this largely makes sense because the high-ranking members of a public company have a fiduciary duty to shareholders and so any political activities of the company must be independent of private political interests. Privately held companies, however, lack such a fiduciary duty. As a result, the lines between the corporate political activity and personal political activity are more blurred. If high-ranking members of private companies are politically active on behalf of their companies, it is functionally similar to the company conducting political activity, but would not be recorded as such. This skews the data against recording the political actions of privately owned companies. Furthermore, the way this thesis defined “fossil fuel company” was narrow. If a broader range of companies were examined, it is very possible the data trends would have been different. Moreover, compared to how long fossil fuel companies have been aware of climate change as an issue and, more importantly, been active in promoting climate change denialism, this work looked at a fairly small window of time. It is possible that if a broader amount of time was examined, the data trends would look different.

FUTURE WORK
Limitations aside, this work establishes several possibilities for future research. Future research should study these same types of political strategies using a broader window of time, segment timeframes differently, or use a more liberal definition of “fossil fuel company” to see how these adaptations affect findings. Additionally, this paper focused only on the prevalence of strategy use; if a company donated $500 once to a climate change-denying legislator it was counted as the same amount of political activity as a company who donated $500 weekly to the same legislator. In other words, political activity was not weighted by the resources it took from the company nor the frequency of its use by that specific company. This consideration was made both to “level the playing field” between fossil fuel companies with more resources than others and because of the difficulty in discerning the financial costs of various political tactics. For the goals of this research, a fossil fuel company’s decision to engage in political strategy should not be negated because of the amount of resources they put into that political activity. Future research should examine both the frequency of uses and resource allocations of various political strategies by the fossil fuel industry.

CONCLUSION
For companies in regulated industries, there is a significant incentive to become politically active and work to ensure that the government passes legislation that benefits, or at least does not harm, the company. Climate change poses a significant threat to the survivability of earth and any attempts to keep the planet hospitable into the distant future will require the elimination or severe reduction of greenhouse gas emissions. For companies whose business is dependent on the continued consumption of fossil fuels, however, efforts to mitigate climate change are detrimental to their profits and therefore, harmful to the interests of their shareholders. As a result, it is in the best interest of fossil fuel companies to delay any actions taken towards mitigating climate change. One effective way to prevent people from trying to solve the climate crisis is to convince them that no such crisis exists. In other words, promoting climate change denial helps fossil fuel companies protect their business interests. This thesis sought to understand which political strategies American fossil fuel companies, as an industry, employed while attempting to influence United States climate change-related legislation in their favor.

The results of this study suggest that fossil fuel companies employ every type of political strategy available to them in order to promote climate change denial. That being said, not all fossil fuel companies are politically active, and not all political strategies are employed with the same frequency. Still, statistical analysis shows that these politically active companies employ these strategies deliberately, showing that fos-
Sil fuel companies are strategic political actors who employ political strategies deliberately to accomplish their goals. On a broader scale, this paper bridges the divide between academic work on corporate political activity and real-life political events. It begins to show how changes in the political landscape of an issue impact how corporations deploy political strategies to deal with them. Moreover, this work studies the big picture of corporate political strategy, as opposed to merely focusing on one aspect of corporate political activity. This study provides another piece of evidence for an increasingly widely known fact: climate change inaction is not the result of ignorant laypeople unable to understand science; rather, it is the result of a coordinated political strategy by fossil fuel companies whose financial interests would be threatened by any significant effort to reduce carbon emissions.

Endnotes

7 Naomi Klein, This Changes Everything: Capitalism vs. the Climate, (Simon & Schuster Paperbacks, 2015), 141-51.
8 “New Pole: Most Americans Want the Government to Combat Climate Change; Some Willing to Pay a High Amount,” Energy Policy Institute at the University of Chicago.
10 Klein, This Changes Everything: Capitalism vs. the Climate, 35; Worland, “Climate Change Used to Be a Bipartisan Issue. Here’s What Changed.”
11 Klein, This Changes Everything: Capitalism vs. the Climate, 35; Worland, “Climate Change Used to Be a Bipartisan Issue. Here’s What Changed.”
12 Franta, “Shell and Exxon’s Secret 1980s Climate Change Warnings.”
24 Klein, *This Changes Everything: Capitalism vs. the Climate*, 23.
25 Ibid, 22.
32 Davenport and Lipton, “How G.O.P. Leaders Came to View Climate Change as Fake Science.”
34 Davenport and Lipton, “How G.O.P. Leaders Came to View Climate Change as Fake Science.”
35 Ibid.
37 Davenport and Lipton, “How G.O.P. Leaders Came to View Climate Change as Fake Science.”
38 Ibid.
39 Ibid.
40 Ibid; Clare Foran and Lauren Fox, “Pelosi Plans to Push for Revival of Climate Change Committee in New Congress,” CNN, November 8, 2018.
42 Ibid.
44 Davenport and Lipton, “How G.O.P. Leaders Came to View Climate Change as Fake Science.”
49 Ibid, 828.
50 Ibid, 828-29.
51 Ibid, 829-30.
52 Ibid, 827-29.
54 Ibid, 829.
55 Ibid, 830.
56 Ibid, 828.
57 Ibid, 831.
60 Ibid, 830.
61 Ibid, 827.

Ibid, 114.

Ibid.

Ibid.


Ibid, 113.


Ibid, 117.


Ibid, 569-70.

Ibid, 570.

Ibid, 571.

Ibid, 569.

Ibid, 570-72.

John Dryzek, Richard B. Norgaard, and David Schlosberg, Oxford Handbook of Climate Change and Society (Oxford University Press, 2014), 146. Klein, This Changes Everything: Capitalism vs. the Climate, 44-46.


Klein, This Changes Everything: Capitalism vs. the Climate, 149.


Ibid, 901.

Ibid, 905.

Ibid.

Ibid.

Ibid, 904.


Ibid, 70.

Ibid, 72.

Ibid, 70-73.


Ibid, 292.

Ibid, 293.

Ibid, 292.

Ibid, 289.


Ibid, 41.

Ibid, 33.

Ibid, 36.

Ibid.

Ibid, 37.

Ibid.

Ibid, 37.

Ibid.


Funk, “Beyond Nonmarket Strategy: Market
Actions as Corporate Political Activity,” 38.
117  De Figuiredo, “Integrated Political Strategy,”
19; Funk, “Beyond Nonmarket Strategy: Market Ac-
tions as Corporate Political Activity,” 38-41.
118  Klein, This Changes Everything: Capitalism
vs. the Climate, 35-36.
119  Worland, “Climate Change Used to Be a Bi-
partisan Issue. Here’s What Changed.”
120  Ibid.
121  Klein, This Changes Everything: Capitalism
vs. the Climate, 35-38.
122  Weidenbaum, “Public Policy: No Longer a
Spectator Sport for Business,” 11-14.
123  Gaither and Gaither, “Marketplace Advocacy
by the U.S. Fossil Fuel Industries: Issues of Represen-
tation and Environmental Discourse,” 587-90.
124  Hansen et al, “Collective Action, Pluralism,
and the Legitimacy Tariff: Corporate Activity or Inac-
tivity in Politics,” 421.
125  Ibid.
126  Brown et al, “The Benefits of a Relational Ap-
proach to Corporate Political Activity: Evidence from
Political Contributions to Tax Policymakers,” 72;
Peo-
bles, “Campaign Finance and Policymaking: PACs,
Campaign Contributions, and Interest Group Influence
in Congress,” 906.
127  De Figuiredo, “Integrated Political Strategy,”
19; Funk, “Beyond Nonmarket Strategy: Market Ac-
tions as Corporate Political Activity,” 33.
128  Funk, “Beyond Nonmarket Strategy: Market Ac-
tions as Corporate Political Activity,” 32.
129  Davenport and Lipton, “How G.O.P. Leaders
Came to View. Climate Change as Fake Science.”
130  Pew Research Center, “More Say There Is
Solid Evidence of Global Warming.”
131  Davenport and Lipton, “How G.O.P. Leaders
Came to View. Climate Change as Fake Science.”
132  Broder, “House Passes Bill to Address Threat
of Climate Change.”
133  Davenport and Lipton, “How G.O.P. Leaders
Came to View. Climate Change as Fake Science.”
134  Ibid.
135  Pew Research Center, “More Say There Is
Solid Evidence of Global Warming.”
136  Davenport and Lipton, “How G.O.P. Leaders
Came to View. Climate Change as Fake Science.”
137  Ibid.
138  Ibid.
139  Francis Choi, “Coalition for Responsible
Regulation v. EPA: An Analysis of Judicial Deference
and Regulatory Discretion,” Ecology Law Quarterly
(2013): 525.
140  Ibid.
142  Pew Research Center, “More Say There Is
Solid Evidence of Global Warming.”
143  Department of Energy, President’s Climate Ac-
tion Plan (2013); Gaither and Gaither, “Marketplace
Advocacy by the U.S. Fossil Fuel Industries: Issues of Represen-
tation and Environmental Discourse,” 586.
144  Davenport and Lipton, “How G.O.P. Leaders
Came to View. Climate Change as Fake Science”;
Utility Air Regulatory Group v. Environmental Protec-
145  Davenport and Lipton, “How G.O.P. Leaders
Came to View. Climate Change as Fake Science.”
146  Park, “6 Obama Climate Policies That Trump
Orders Change.”
147  Park, “6 Obama Climate Policies That Trump
Orders Change.”;
148  Jared Keller, “Oil and Gas Ties Run Deep in
the Trump Administration,” Pacific Standard, January
5, 2018; Milman, “No Shame: How the Trump Admin-
istration Granted Big Oil’s Wishlist.”
149  Funk, “Beyond Nonmarket Strategy: Market
Actions as Corporate Political Activity,” 36.
150  2009 Congressional Record, Vol. 155, Page
H199.
151  “Yale Climate Opinion Maps 2018,” Yale Pro-
gram on Climate Change Communication (2018).
152  “Yale Climate Opinion Maps - U.S. 2016,” Yale
Program on Climate Change Communication (2016).
153  Klein, This Changes Everything: Capitalism
vs. the Climate, 12-13; Kate Grist, “How to Change
the Minds of Climate Deniers,” The Guardian, Febru-
ary 3, 2019.
154  Weidenbaum, “Public Policy: No Longer a
Spectator Sport for Business,” 9.
155  Klein, This Changes Everything: Capitalism
vs. the Climate, Introduction.

Bibliography

Ard, Kerry, Nick Garcia, and Paige Kelly. “Another
Avenue of Action: An Examination of
Climate Change Counterrunvention Industries’ Use of PAC Donations and Their Relationship to Congressional Voting over Time.”


Bonardi, Jean-Philippe, and Gerald D. Keim.


Brooks, J. C., A. C. Cameron, and C. A. Carter.


Gaither, Barbara Miller, and T. Kenn Gaither.


Island Journal 8, no. 2 (1993).

text here...


Wolfsfeld, Gadi. “Political Actors.” Informing Conflict Prevention, Response and Resolution.


DANGEROUS DECEPTION: USING LITIGATION TO MINIMIZE THE MEDICAL MISINFORMATION DISSEMINATED BY CRISIS PREGNANCY CENTERS

HOLLY SCRUGHAM, POMONA COLLEGE CLASS OF 2019

ABSTRACT
This paper examines litigation based strategies used to regulate crisis pregnancy centers (CPCs). CPCs are facilities run by anti-choice organizations designed to dissuade pregnant people from getting an abortion. Until recently, much of the academic scholarship focusing on regulating CPCs has looked at attempts by state and municipal governments to require CPCs post signs in their waiting rooms disclosing various pieces of information, like the fact that they are not licensed to practice medicine or that they do not provide abortion. CPCs have generally been successful in challenging these laws in the courts, most notably in the 2018 Supreme Court case NIFLA v Becerra. This paper argues that the door to regulating CPCs through these mandated disclosure ordinances has closed and pro-choice advocates should explore new and innovative legal strategies to achieve further regulation. In particular, this paper analyzes a recent argument proposing raising a tort claim of battery against CPCs. Essentially, if it could be established that clients consent to being touched by CPCs under the false premise that the CPC is a bona fide health clinic, then the consent was made under false pretenses and is therefore invalid. This paper analyzes that argument in a broader context of political science theory on making social change through litigation.

In 2002, a woman named Cherisse opened the Chicago Yellow Pages to look for an abortion clinic. She made an appointment, but when she arrived, she wasn’t greeted by a nurse or doctor who would provide an abortion. Rather, the staff member who met Cherisse showed her a graphic video of a Dilation and Evacuation (D&E) abortion and tried to convince her to carry her pregnancy to term.1

The center that Cherisse visited was not an abortion clinic, but a facility called a crisis pregnancy center (CPC).1 CPCs are anti-choice facilities designed to dissuade pregnant people from getting an abortion by creating the appearance of a genuine medical center and spreading lies and misinformation about abortion.2,3 They are a part of the anti-choice movement’s structured, systematic attempt to prevent abortion.

The first CPC sent Cherisse to another CPC to get an ultrasound. There, a staff member told her that “[i]f you have an abortion now, you’ll rupture your uterus and won’t be able to have children in the future.”3 Cherisse had no idea what was true, and did not want to have future regrets about not being able to have children. She decided not to get an abortion. She later lost her job and could not afford rent, food, utilities, or healthcare.4

Cherisse’s story is not unique. Though we do not know how many pregnant people make the decision not to get an abortion based off of misinformation given to them by a CPC, we do know that the practices of the CPCs Cherisse visited are standard. While CPCs do employ a variety of tactics and there is not one singular CPC experience, scholarship on CPCs depicts a fairly standard scene.

It likely starts with a missed period, a positive portant and gives insight into the broader, religiously conservative agenda of CPCs, but the scope of this paper is limited to how the relationship between CPCs and abortion.

1 These facilities are sometimes referred to as pregnancy resource centers or limited service pregnancy centers. I will be referring to them as crisis pregnancy centers.
2 It has also been documented that CPCs give medically inaccurate information about birth control and contraception, and promulgate an abstinence-only viewpoint in their counseling sessions. This is in-
pregnancy test, and a Google search. Searches like “pregnancy resources near me” contain results for CPCs on the first page of the search results (see Figure 1). If called, the CPC will probably encourage the pregnant person to come into the clinic, likely avoiding divulging any information that would identify them as staunchly anti-abortion. The clinic will be nondescript, maybe in a strip mall or office building (See Figure 2). Once the pregnant person arrives at a clinic, they may be greeted by a volunteer in a white lab coat and asked to fill out medical paperwork. The client will be taken inside the clinic, possibly have their blood pressure taken, and will be given a pregnancy test. The client may be given an abdominal or transvaginal ultrasound, depending on the equipment the CPC has. Then the client will be “counseled” by a CPC volunteer or staff member, a conversation in which the CPC staffer will try to convince the client not to have an abortion.

Pro-choice advocates have long recognized that CPCs consistently use deception and give medically inaccurate information about abortion to their clients, and have made various attempts to regulate them. Part I of this paper explores the research suggesting that CPCs across the country consistently use deception and give medically inaccurate information about abortion to their clients. Many small-scale studies conducted by various research institutions have investigated the extent to which CPCs give medically inaccurate information. NARAL Pro-Choice America is one of the U.S.’s largest pro-choice organizations and conducted the majority of the studies utilized in this paper. With a few exceptions, these studies have generally each investigated only 15-25 CPCs in individual states, yet they provide strong anecdotal evidence of misinformation and deception. This paper helps fill that research gap; by analyzing seven individual studies and reports and compiling their data, I was able to make conclusions about the tendencies of 190 CPCs in geographically diverse states across the country, which more strongly suggests that deceptive practices are widespread among CPCs and are not isolated incidents.

LITERATURE REVIEW
Several scholars have studied the various attempts made by pro-choice advocates to counter the propagation of deceptive medical information. Existing inquiries largely focus on mandatory disclosure ordinances, which are laws requiring that CPCs disclose various pieces of information by posting notices in their waiting rooms or in their entryways, like the fact that they are not licensed medical clinics or that they do not
provide or refer for abortion. CPCs have successfully challenged the disclosure laws in court on the grounds that their free speech was being abridged. Most notably, in the 2018 case *NIFLA v. Becerra*, the Supreme Court struck down the California Reproductive FACT Act, a law requiring CPCs to post notices stating that California offered low cost abortion services, as well as whether or not the CPC was licensed.12 This decision falls into a larger trend of court hostility to preserving abortion access. By examining six disclosure ordinances, particularly the FACT act, as well as the court’s history of ruling against abortion access, this paper will show in Part I that the door to regulating CPCs by passing mandated disclosure laws has effectively closed.

Part II of this paper seeks to answer this question: when considering that mandated disclosure notices are no longer legal viable regulation options, what is the most effective litigation-based strategy that pro-choice advocates can use to minimize the ability of CPCs to promulgate misinformation? I will argue that advocates should work on creative and innovative legal strategies that are more resistant to anti-abortion judges. One of the most promising solutions that this paper will explore is an innovative tort strategy proposed in 2018 by Teneille Brown. She suggests bringing a claim of battery against CPCs who touch clients under the false premises that they are a bona fide medical clinic.13 This paper analyzes Brown’s proposal by situating it into the larger field of political science scholarship that examines how progressive advocates can make social change through litigation. Alison Gash’s 2015 book *Below the Radar: How Silence Can Save Civil Rights* proposes a novel strategy of change-making litigation – one that Brown’s tort approach demonstrates. Alison Gash argues that in certain cases, the degree to which advocacy is “visible” can seriously hinder its efficacy – high visibility can create debilitating backlash, while low visibility can circumvent that hostility.14 This paper will argue that the high visibility of abortion is one of the reasons that mandated disclosures regarding abortion have failed in the courts. Consequently, regulating CPCs with tort claims – which are less visibly about abortion – is one of the biggest strengths of Brown’s proposal.

Brown’s proposal is promising, but lacks an in-depth examination of its own drawbacks. In this paper, I will use Marc Galanter’s 1976 work “Why the ‘Haves’ Come Out Ahead: Speculations on the Limits of Legal Change” to illuminate some of the limitations associated with attempting to make social change through litigation, particularly when pitting a well-financed, politically and legally savvy organization like a CPC against a single plaintiff. Galanter proposes that there are two kinds of actors in the legal system: those who are repeatedly involved in litigation, like insurance companies, and those who are very rarely involved in litigation. For the former, the stakes of each case tend to be much lower, and they are able to strategically game the legal system in order to gradually shape the law in their favor. The latter, on the other hand, is highly concerned with the outcome of their individual case, and is not likely to be concerned about the ultimate state of the law.15 Galanter’s work is helpful in conceptualizing the “games” that each actor in Brown’s proposal would be playing: the CPC would be playing the long game, highly concerned by the outcome of the law, while the individual plaintiff who has experienced battery would theoretically be unconcerned with the outcome of the law and solely interested in receiving compensation in the form of damages. Galanter’s work highlights one major limitation of Brown’s proposal: CPCs are already in a position of legal power over individual plaintiffs. This could lead to a situation in which a CPC proposes an enticing settlement offer to a plaintiff instead of risking losing a case and creating law that would ultimately hurt the CPC.

While it is true that Brown’s proposed plaintiff would not share all of the same characteristics as Galanter’s traditional individual plaintiff, as Brown’s plaintiff would also likely be interested in establishing legal precedent, Brown also does not examine the ethical implications of using a “test-case” plaintiff in this way. Dennis Katz’s 1971 article “The Public’s Interest in the Ethics of the Public Interest Lawyer” illuminates a limitation of Brown’s proposal, noting that cases that prioritize changing law above the individual plaintiff are inherently contradictory to the fundamental purpose of attorneys, which is to zealously represent the interests of the client.16 This paper does not argue that these potential drawbacks mean Brown’s proposal isn’t worth pursuing, but works to ensure that all aspects of the strategy, both advantages and disadvantages, are being considered. Ultimately, I argue that the lower-visibility aspects of Brown’s strategy make it extremely promising and potentially powerful, but advocates should pay careful attention to the ethical implications of litigating with a test-case plaintiff when employing it.
METHODOLOGY
This paper conducts a two-part examination of the issue of CPC regulation. Part I explores the research suggesting that CPCs across the country consistently use deception and give medically inaccurate information about abortion to their clients. I analyze seven distinct studies: a 2006 Congressional study called the Waxman Report; a 2012 study by the Department of Obstetrics and Gynecology at the University of North Carolina Chapel Hill; and investigations from the following state NARAL affiliates: Montana, New York, Ohio, North Carolina, and Maryland. Each of these studies investigated about 15-25 CPCs. The methods of these studies were very similar: researchers posed as pregnant women and either called or conducted a “secret-shopper” investigation, in which the researchers visited the CPC posing as someone who thought they were pregnant. When asserting that CPCs spread medical misinformation, many scholars point to one or two of these individual studies, meaning their assertion rests on information gathered from less than thirty CPCs that may all be in the same state or region. In order to attempt to make a broader claim about the general practices of CPCs, I analyzed deception trends that occurred across the studies and compiled data on a total of 190 clinics. Specifically, I looked at how many CPCs investigated across the seven studies were reported to disseminate three general categories of misinformation: (1) a link between abortion and breast cancer (2) the effects of abortion on future fertility/damage abortion has on the reproductive system and (3) the mental health effects of abortion.

In order to contextualize the need for new and innovative legal strategies to combat CPCs’ deception, Part II of this paper begins by summarizing the repeated failed efforts to regulate CPCs with disclosure ordinances in Baltimore, Montgomery County, MD, New York City, Austin, and San Francisco. I then present a detailed discussion of the Reproductive FACT Act, why it failed in the courts, and why its failure means that fighting for mandating disclosure laws is no longer the best use of the time and resources of pro-choice advocates. I analyze other strategies that pro-choice advocates should pursue to regulate CPCs, focusing particularly on Brown’s tort claim strategy but also including a discussion of using the Establishment Clause as a way to decrease the government funding received by CPCs. In conclusion, this paper adds to the literature on strategies to regulate CPCs by helping to fill the research gap of larger scale studies of the misinformation spread by CPCs and ultimately proposes that pro-choice advocates should employ Brown’s tort claim strategy while bearing carefully in mind its ethical ramifications.

HISTORY AND BACKGROUND
CPCs became prevalent in the United States in the mid to late 1960s in response to the passage of more liberal abortion laws in states across the country. Many attribute the founding of the modern CPC model to a man named Robert Pearson, who opened one of the first CPCs in 1967 in Hawaii. Soon after, Pearson created the Pearson Foundation, an organization based in St. Louis designed to help other groups found crisis pregnancy centers. The foundation provided free and low cost training materials, including pamphlets, video equipment, pregnancy tests, and slide shows, including one titled “Caring,” a nearly 30-minute presentation that “includes many pictures of bloody fetuses in waste cans and one of a gurney carrying a woman who is apparently dead and is covered by a sheet. It ends by comparing abortion to the final solution.” Pearson also created a manual, How to Start and Operate Your Own Pro-Life Outreach Crisis Pregnancy Center, which government officials have called “a 93-page guidebook of ways to mislead consumers.” The manual details numerous ways in which CPC staff can deceive clients into thinking they provide abortion services, such as “there is nothing wrong or dishonest if you don’t want to answer a question that may reveal your pro-life position by changing the caller’s train of thought by asking a question in return.” The manual also instructs staff to answer the question “Are you a pro-life center?” with “We are a pregnancy testing center...What is pro-life?”

Today, the majority of CPCs is affiliated with one or more religiously-affiliated umbrella organizations that follow a model similar to the one espoused by Pearson. The main CPC umbrella organizations in the U.S. are Care Net, Heartbeat International, Birthright International, and the National Institute of Family and Life Advocates, or “NIFLA.”
Table 1

<table>
<thead>
<tr>
<th>Umbrella Organization</th>
<th>Year Established</th>
<th># of CPCs Worldwide</th>
<th>Operates brick-and-mortar facilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Care Net</td>
<td>1975</td>
<td>1,100</td>
<td>Yes</td>
</tr>
<tr>
<td>Heartbeat International</td>
<td>1971</td>
<td>2,500</td>
<td>Yes</td>
</tr>
<tr>
<td>Birthright International</td>
<td>1968</td>
<td>Not listed on website</td>
<td>Yes</td>
</tr>
<tr>
<td>NIFLA</td>
<td>1993</td>
<td>1,400</td>
<td>No – just legal support</td>
</tr>
</tbody>
</table>

While the primary purpose of CPCs is to dissuade clients from getting abortions, many also heavily promote an abstinence-only sexual education agenda by providing misinformation about contraception and encouraging abstinence until marriage. This funding has continued throughout the years, particularly under the Trump administration. In March 2019, the Trump administration announced that the Department of Health and Human Services would allocate $5.1 million to the Obria Group, a California based crisis pregnancy center network, in order for them to expand their services in three California counties. CPCs also receive funding from specialty “Choose Life” vanity license plates that are currently available for purchase for an additional $25 to $70 in thirty-two states. Of these states, seventeen donate proceeds from the vanity plate sales to directly to CPCs, and ten states explicitly prohibit the money raised from the Choose Life plates from going to any agency or organization that provides abortion counseling, referral, or procedure.

DECEPTION OUTSIDE THE CPC ADVERTISING

CPCs execute aggressive and ubiquitous advertising campaigns across a variety of media. They rely heavily on deception tactics in order to bring what they call “abortion-minded” or “abortion-vulnerable” clients to their facilities under the guise of being a comprehensive medical clinic into their facilities. “Choose Life Marketing,” a pro-life PR firm that works with CPCs on their advertising, states on its website that it “serves pregnancy centers with a set of marketing services developed to help them reach more abortion-minded women.” In particular, CPCs target “abortion-minded” women who come from more vulnerable populations; CPCs disproportionately advertise and are located in low-income neighborhoods and near high schools and colleges.

Advertisements for CPCs are often vague, using words like “options” and “choice.” This vague language is typical.
language also appears in the names of clinics: “Options” and “Choices” are very common CPC names.\(^{39}\) Furthermore, CPCs buy keyword-based advertising space on Google on words like “abortion” and post ads that mimic the content of actual abortion facilities.\(^{40}\) When I Google “pregnancy center,” the first sponsored ad is for a CPC with the heading “Pregnancy Resources Center | Free Medical Services,” and the sub-heading is “Learn More. We Can Help. Talk To Us Now. Free and Confidential Services. 24-Hour Help Line. Free Counseling. Confidential Appointments. Services: Services: Free Pregnancy Testing, Free Prenatal Consult, Free STD/HIV screening, 24-hour help line” (See Figure 3).

Figure 3: ”Pregnancy Center” Google search result (3/20/19)\(^{41}\)

When I Google “abortion clinic Los Angeles,” the first four sponsored ads do lead to websites of clinics that offer abortion, but the fifth ad is titled L.A. Women’s Clinic and leads to the website “yourabortionchoice.org. The sub-heading of the ad is “Medically accurate abortion info. Free, Confidential & No Judgment” (See Figure 4).

Figure 4: “Abortion clinic Los Angeles” Google search result, 3/20/19\(^{42}\)

The website is fairly bare and does not link to a broader CPC network, like Birthright or Care Net, but it still appears to be a CPC, advertising free pregnancy tests, “pregnancy options education” and “pre-pregnancy termination counseling.”\(^{43}\) The website reads, “You have the right to know all the options — as well as all potential risks to you — before you decide what you should do regarding an unplanned pregnancy.”\(^{44}\) It also advertises “medically accurate abortion information” and says “we discuss or provide information on Abortion Cost, Plan B/Morning After Pill/RU486, and abortion alternatives.”\(^{45}\) This language reflects the broad strategy of CPCs: stay vague, positive, and avoid concrete statements about abortion. For example, NARAL reports that 75 percent of CPCs surveyed in New York City do not disclose on their websites that they are against abortion, while 69 percent of clinics that NARAL surveyed in California advertise “unbiased” counseling as one of their services.\(^{46}\) Of the main umbrella networks, Heartbeat and Care Net both disclose on their websites that they do not provide or refer for abortion.\(^{47}\) Birthright does not make this disclosure; in fact, it does not mention the word “abortion” a single time on its website.\(^{48}\) In this way, CPCs appear to have evolved from Pearson’s model, which overtly encouraged hiding the fact that the CPC did not provide abortion.

NARAL successfully campaigned in 2014 for Google and Yahoo to remove overtly deceptive ads by CPCs. They gave a report to these companies containing “extensive documentation” of CPCs’ false advisements and later reported that Google and Yahoo had removed over two-thirds of the advertisements.\(^{49}\) Although this is undoubtedly a step in the right direction, and advocates should continue to monitor and report to search engines’ deceptive advertising, it is clear that CPCs still engage in deceptive advertising. One significant limitation to the effort to stop deceptive advertisements is their proliferation – it is nearly impossible for NARAL and other advocates to find and report every single one. These organizations are well-financed and spend heavily on advertisements: NARAL reported in 2015 that “Care Net and Heartbeat International spend more than $18,000 per month on pay-per-click advertising campaigns that target women searching for abortion providers and bring them to their websites.”\(^{50}\)

**CO-LOCATION STRATEGIES**

In 2011, a CPC called “Problem Pregnancy” opened a clinic a few doors down from a Planned Parenthood, on the same floor of the same building, even using the same acronym (PP).\(^{51}\) The Planned Parenthood moved its location, but Problem Pregnancy relocated as well, 5

This CPC also conflates the Morning After Pill, which is not an abortifacient, with RU486, the “abortion pill”, which is a pill that does in fact induce abortion.
directly across the street from the new Planned Parenthood location. This is another tactic to attract “abortion-minded” women as they open centers near bona fide reproductive health clinics in the hopes that patients will accidentally walk into the CPC instead of the comprehensive health clinic. For example, NARAL reported in 2014 that over half of abortion providers in Massachusetts are located within ten miles of a CPC and, in Boston, within two miles. In Ohio, as of 2013, 20% (3 out of 15) of abortion clinics were located on the same block as a CPC, and one even had two separate CPCs within one block.

CPCs also purposefully place themselves in strategic locations to target what they consider the most “vulnerable” populations: low-income, people of color, and high school and college students. In 2003, Care Net launched its “Urban Initiative,” which states that its primary goal is reaching African Americans and Hispanic women by opening more CPCs in communities with higher concentrations of these demographics.

DECEPTION INSIDE THE CPC

Once a person is inside a CPC, the center engages in a variety of deceptive and misleading tactics in order to dissuade or prevent them from obtaining an abortion. Volunteers, who are usually not medical practitioners, don white lab coats as clients fill out official-looking paperwork. One of the most influential tools CPCs use to create the appearance of a medical clinic is the ultrasound, which has become increasingly popular among CPCs. NIFLA states on their website that it

... recognized the importance of using ultrasound in a pregnancy center setting for reaching abortion-minded women more than two decades ago, and has been pioneering the way in which the pro-life movement used this important tool ever since. Ultrasound offers a window to the womb which can impact a woman’s decision to choose life by more than 80 percent.

The anti-choice right wing organization Family Research Council (FRC) released a report in July 2010 about the importance of ultrasounds to the movement. These findings have not been confirmed, but they help demonstrate the importance CPCs place on ultrasounds. Some of the findings reported by FRC from various sources include:

• women planning to abort their child were 2 – 3 ½ times more likely to keep their children after viewing their ultrasounds.
• women at high risk of having abortion who viewed ultrasounds were only 30% likely to continue with the procedure, compared to 75% of women who did not view ultrasounds.
• “abortion-minded” women decide to keep their babies after seeing ultrasound images.
• 90% of women who see their baby by ultrasound choose life.
• “ninety-eight percent of women who have ultrasounds choose to carry to term.”

The ultrasound is one of the many tools utilized by CPCs to create the sense that their facility is a comprehensive health clinic. After gaining this trust from clients, CPCs routinely propagate misinformation about abortion.

Many studies have documented that CPCs give false or misleading information about abortion. One of the first investigative efforts into CPCs was conducted in 2006, when Rep. Henry A. Waxman (D-CA) released a comprehensive report on CPCs titled False and Misleading Health Information Provided by Federally Funded Pregnancy Resource Centers. In this study, researchers called twenty-three CPCs, posing as a 17-year-old girl who thought she was pregnant and was trying to decide whether or not to have an abortion. They found clear evidence that CPCs promulgated medically inaccurate information that fell into three general categories: (1) a link between abortion and breast cancer, (2) the effects of abortion on future fertility and the damage abortion has on the reproductive system, and (3) the mental health effects of abortion. Since the Waxman Report, many similar studies have been conducted, largely by state affiliates of NARAL Pro-Choice America. Their methodologies have been similar: researchers posed as pregnant women and either called or conducted a “secret-shop” investigation, in which researchers visited a CPC posing as someone who thought they were pregnant. The findings of these studies are consistent: CPCs very frequently give medically inaccurate information.

However, like the Waxman Report, the individual studies by NARAL tend to have small sample sizes of between 15 and 25. To determine whether similar findings occurred across these studies, I analyzed several reports and compiled data on their findings in each of these three categories. The data that follows are based on findings from Waxman Report; a 2012
study by the Department of Obstetrics and Gynecology at the University of North Carolina Chapel Hill; and investigations from NARAL state affiliates in Montana, New York, Ohio, North Carolina, and Maryland. These data are primarily based on secret-shopper surveys, with some data coming from phone calls. I compiled data on 190 clinics.

Even though the medical community agrees that there is no link between abortion and an increased risk of breast cancer, 44% (83) of CPCs studied stated that abortion would increase the risk of breast cancer. The Waxman Report says that one center told callers that “all abortion causes an increased risk of breast cancer in later years,” and another stated that “research shows a ‘far greater risk’ of breast cancer after an abortion” and that “the risk of breast cancer increased by as much as 80% following an abortion.”

Additionally, 59% (113) of the CPCs stated that abortion was linked to or would increase the likelihood of infertility, post-abortion scarring/injury, or some combination thereof. One CPC investigated in the Waxman Report told the caller that “damage from abortion could lead to ‘many miscarriages’ or to ‘permanent damage’ so that you ‘wouldn’t be able to carry.’” According to a 1990 study called “The Effect of Pregnancy Termination on Future Reproduction” published in Bailliere’s Clinical Obstetrics and Gynecology, abortions that happen in the first trimester of pregnancy—which make up about 92% of all abortions—are not linked to future infertility, ectopic pregnancy, miscarriage, premature delivery, or birth of babies with a low birth weight. There is some risk of future premature delivery or low-weight birth following second-trimester abortions that are performed using dilation and evacuation, but CPCs were not found to distinguish between abortions performed at different stages of pregnancy. Furthermore, abortion is one of the safest medical procedures: with a less than .05% risk of major complications, abortion is safer than a tonsillectomy.

67 percent of the CPCs studied gave misinformation about the psychological effects of abortion. Pro-life groups assert the existence of “post-abortion syndrome,” a psychological condition allegedly similar to post-traumatic stress disorder. According to NARAL, a Missouri CPC volunteer said that “you might feel sad every year around the time the baby was supposed to be born, or that you might wince or react negatively when you hear a vacuum or dentist’s drill.” The Waxman Report states that “numerous methodologically sound studies” have not been able to confirm the existence of such a condition. In fact, research shows that “significant psychological stress after an abortion is no more common than after birth” and that relief is the predominant emotion experienced after an abortion. A longitudinal study of 843 women who received abortions between 2008 and 2010 at thirty U.S. abortion clinics compared the emotions of three groups of women: (1) those who had received an abortion right before the abortion clinic’s gestational limit, (2) those who were denied an abortion because their pregnancy was slightly too far along, and (3) those who had obtained a first-trimester abortion. The study found that those in the second category who were denied an abortion felt more regret and anger and less relief and happiness than those in the first category. The study found that 95 percent of those who obtained an abortion felt it was the right decision.

One of the most extreme techniques used by CPCs that researchers have documented is attempting to delay a client’s decision about abortion until it becomes legally impossible. Delay tactics have not been researched as much as the aforementioned types of misinformation, but there have been numerous documented occurrences. CPCs have been found to “suggest the possibility of miscarriage as a reason to avoid making an immediate decision on abortion.” NARAL North Carolina found that 24 percent (16 out of 66) of CPCs studied used this deceptive tool, and a UNC study found that 23 percent (3 out of 19) of CPCs researched stated that “pregnancy carries a substantial risk of miscarriage and that ‘there is plenty of time,’ despite not having specific information on the researcher’s gestational age.” While the American College of Obstetricians and Gynecologists reports that miscarriages happen in about 10 percent of pregnancies, making them far from rare, using the possibility of miscarriage as a tool to delay a decision about abortion is medically irresponsible.
who would perform an abortion after viability.77 An investigator in New York who visited a CPC posing as a woman who was 9.3 weeks pregnant had a similar experience: a staff member told her, “in this country you can get an abortion up to nine months” and that she “had time to think about it.”78 In reality, when this article was written in 2008, abortion in New York was illegal after twenty-four weeks.79

CONSTITUTIONAL LAW CONTEXT
FREE SPEECH
CPCs present a clear and immediate threat to access to comprehensive reproductive healthcare. For this reason, pro-choice advocates have been challenging CPCs since they began to proliferate. Before discussing the most effective strategies for minimizing the potential harm of CPCs, it is important to analyze the measures pro-choice advocates have already taken in an effort to regulate as well as the reasons these efforts have failed, proving that these paths are no longer worth pursuing. Throughout the past decade, pro-choice advocates and lawmakers have tried to regulate CPCs by passing city or county ordinances requiring that CPCs disclose certain pertinent information to clients, usually by posting written notices in their waiting rooms or directly outside the center. These notices were designed to inform clients that these centers are not comprehensive reproductive healthcare centers in order to encourage them to seek care from a licensed medical provider. Prior to California’s Reproductive FACT Act, which I will discuss in more detail in the following chapter, four cities and one county had enacted mandatory disclosure laws: Baltimore; Montgomery County, MD; New York City; Austin; and San Francisco. I will discuss the Baltimore case in detail, as it was the first disclosure ordinance to be litigated and includes many of the overarching elements found in the other ordinance cases. A chart summarizing the other ordinances and the legal events surrounding them will follow.

TYPES OF DISCLOSURES
The kinds of disclosures mandated by these laws can be grouped roughly into three categories: service disclosures, status disclosures, and government messages disclosures.80 Service disclosures mandate that CPCs reveal that they do not provide certain reproductive healthcare services, like abortion, emergency contraception, or birth control.81 Status disclosures require a disclaimer about whether or not there is a licensed medical provider on staff or if the CPC is a licensed medical facility.82 Government message disclosures require publicizing the advice of government officials or agencies, which is generally that the city health department suggests that those who are pregnant seek healthcare from a licensed medical provider or that the city/state offers low-cost reproductive healthcare.83

KEY CONSTITUTIONAL QUESTIONS
Each of these ordinances compelled some kind of speech from the CPC, and courts have generally ruled that any law compelling speech is subject to strict scrutiny. When laws either compel or restrict speech, courts look at them with three types of scrutiny: strict, intermediate, and rational basis review. The first two types of scrutiny, strict and intermediate, are relevant in the discussion about mandated CPC speech ordinances. Strict scrutiny requires that the law be “narrowly tailored” to promote a “compelling government interest” and that “if a less restrictive alternative would serve the government’s purpose, the legislature must use that alternative.”84 When intermediate scrutiny is applied, it means that “both the urgency of the government’s goals and the closeness of the means-end fit need not be go great as in strict scrutiny.”85 In First Amendment cases, the compelled speech of the law must be “substantially related” to the government’s interests, a step down from the “narrowly tailored” requirement of strict scrutiny.86

In 2009, Baltimore became the first city in the country to pass a CPC disclosure mandate. The rule was added to the Baltimore City Health Code for the purpose of “provid[ing] its clients and potential clients with a disclaimer substantially to the effect that the center does not provide or make referral for abortion or birth-control services.”87 Because the mandated disclosure pertained to the services CPCs did not offer, it is a service disclosure. The law called CPCs “limited-service pregnancy centers” and defined them as “any person. . . whose primary purpose is to provide pregnancy-related services [and] provides information about pregnancy-related services [but] does not provide or refer for . . . abortions [or] nondirective and comprehensive birth-control services.”88 It mandated that “at least one ‘easily readable’ sign ‘written in English and Spanish’ must be ‘conspicuously posted in the center’s waiting room or other area where individuals await service.’” The fine for failing to post such disclosure was a $500 misdemeanor fine for each day the sign is not posted.89
The Archbishop of Baltimore, Edwin F. O’Brien, challenged the ordinance in the Maryland District Court in conjunction with the Greater Baltimore Center for Pregnancy Centers, Inc on First Amendment grounds. The court’s task was to decide what level of scrutiny to apply to the mandated-disclosure ordinance and to decide if the ordinance would pass that level of scrutiny. Plaintiffs argued that strict scrutiny needed to be applied because (1) the speech was not commercial in nature and (2) the speech was regulated on the basis of defendants’ ideological and viewpoint disagreement with plaintiffs. Respondents argued that a lower level of scrutiny should be applied to the mandate because it did in fact regulate commercial speech and is “viewpoint-neutral.”

The constitutional law notions of commercial versus non-commercial speech and viewpoint-neutrality have come up repeatedly in legal challenges to various CPC mandated disclosure notices and warrant further analysis. Proponents of the mandated disclosure statutes have tried, largely unsuccessfully, to argue that the CPCs’ speech is commercial. The Court decided in the 1980 case *Central Hudson Gas v. Public Service Commission* that commercial speech should be held to intermediate scrutiny, stating that a law that regulates commercial speech must “directly advance a substantial governmental interest” and is “not more extensive than necessary to serve that interest.” *Central Hudson* also defined commercial speech as “expression related solely to the economic interests of the speaker and its audience.” The Court held in a later case, *Milavetz, Gallop & Milavetz v. United States*, that laws regulating commercial speech are valid in the case that the “disclosure requirements are reasonably related to the State’s interest in preventing deception of customers.” Courts also apply higher scrutiny when they determine that a regulation discriminates on the basis of viewpoint, meaning that the regulation endorses a certain political ideology over another. For example, a law prohibiting speaking out against abortion would be “viewpoint discrimination” because it endorses a pro-choice political viewpoint.

The City of Baltimore argued that the speech in question was commercial because CPCs were providing women “with valuable goods and services” like pregnancy tests, sonograms, and options counseling. The court disagreed with this argument, stating, “the overall purpose of the advertisements, services, and information offered by the CENTER is not to pro-

The services offered by the CPC had no “value in the commercial marketplace” because they were free. Thus, in deciding that the speech was not commercial, the court applied strict scrutiny to the mandated disclosures. The court also agreed with plaintiffs that the law was viewpoint discriminatory, stating that it was formed on “disagreement with plaintiff’s viewpoints on abortion and birth control.” The ordinance failed to pass strict scrutiny and was struck down by the court. The court held that it was not “narrowly tailored,” as it did not provide a “carve-out” provision for CPCs that “do not engage in deceptive practices”—essentially, the law needed to have a caveat that prevented non-deceptive CPCs from being affected by the law. The court also argued that it would have been possible to enact a less restrictive regulation “regarding fraudulent advertising to combat CPCs’ deceptive advertising campaigns, or adopt a content-neutral advertising ordinance that applied to non-commercial entities.”

Similar laws have been enacted and challenged in the courts in Montgomery County, Maryland; Austin, Texas; New York City, New York; and San Francisco, California. The following chart summarizes the key information associated with each ordinance.
Table 2: Speech Ordinances and Legal Challenges

A number of factors listed in the New York City and San Francisco statutes dictate what constitutes the appearance of a licensed medical facility. If a CPC has two or more of the following characteristics, it can be considered to have the appearance of a medical clinic:

(a) offers pregnancy testing and/or pregnancy diagnosis; (b) has staff or volunteers who wear medical attire or uniforms; (c) contains one or more examination tables; (d) contains a private or semi-private room or area containing medical supplies and/or medical instruments; (e) has staff or volunteers collect health insurance information from clients; and (f) is located on the same premises as a licensed medical facility or provide or shares facility space with a licensed medical provider.

The San Francisco ordinance is unique in that the mandatory disclosure notices were only required if the CPC was engaging in false, misleading, or deceptive advertising. If the CPC is advertising in a way that is “untrue or misleading, whether by statement or omission,” they would have ten days to “cure” the problematic advertising; if they do not, it would trigger the requirement to post the disclosure notices.

**Table 2: Speech Ordinances and Legal Challenges**

<table>
<thead>
<tr>
<th>San Francisco, 2011</th>
<th>#1 Status disclosure</th>
<th>#2 Services disclosure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>On a conditional basis if advertising is deceptive</td>
<td></td>
</tr>
<tr>
<td>#1: Whether or not center has a licensed doctor or nurse on staff</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#2: Whether or not center provides abortion, emergency contraception, or referrals for these services</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Ninth Circuit (2015): Applies intermediate scrutiny because the ordinance focuses on false and misleading commercial speech. Ordinance is deemed constitutional.** In 2018, the Supreme Court denied to hear an appeal from petitioners.

**REPRODUCTIVE FACT ACT**

In 2018, California’s Reproductive FACT Act became the first CPC mandated disclosure notice to be litigated at the U.S. Supreme Court. The Act, mandating that pregnancy-related clinics post signs notifying that (1) California offers low-cost reproductive health services and (2) disclosing the licensing status of the clinic, was signed into law in October 2015. Several CPCs immediately challenged the Act on the grounds that it violated their freedom of speech and freedom of religion. The CPCs’ motion for an injunction was denied by federal courts in California’s Eastern and Southern Districts in 2015, and in 2016, the Ninth Circuit Court upheld the decision of the District Courts. The Supreme Court overturned the Ninth Circuit’s decision in June of 2018.

California’s ordinance was different from its predecessors in two significant ways. It mandated two different messages depending on whether the facility was licensed, the first of which actually applied both to CPCs and bona fide reproductive health clinics. Provision #1 applied to “licensed covered facilities,” which it defined as a licensed clinic “whose primary purpose is providing family planning or pregnancy related services,” and offers at least two of the following: obstetric ultrasounds, sonograms, prenatal care, counseling about contraception, pregnancy testing, pregnancy options counseling, abortion services, or staff or volunteers who gather health information from clients. This definition adds two factors to the list established by the NYC and San Francisco ordinances: contraception counseling and abortion services. This means that the first provision of the ordinance would apply not only to CPCs, but to comprehensive reproductive health clinics as well. Lawmakers decided to broaden the applicability of the statute in this way to help protect it against the claim that the regulation was targeting facilities affiliated with a religious ideology. Facilities in this category were required to disseminate what can be categorized as a government message disclosure stating

---

7 It is somewhat puzzling that the Court declined to hear petitioner’s appeal on the SF ordinance two days after it struck down a very similar law in the NI-FLA case. It is possible that the Court thought that this would create bad law and raise questions they did not want to answer. Interestingly, this ordinance contains the kind of “carve-out” option that the court wanted to see in the Baltimore ordinance (the option for non-deceptive CPCs to opt out of the ordinance). This is a question that should be explored in further research. “First Resort, Inc. v. Herrera,” SCOTUSblog, accessed April 25, 2019, https://www.scotusblog.com/case-files/cases/first-resort-inc-v-herrera/.

8 The State Assembly was influenced by a report called “Pregnancy Resource Centers: Ensuring Access and Accuracy of Information” conducted by the Public Law Research Institute at UC Hastings College of Law. This report argued that the biggest issue of past disclosure ordinances was that they targeted religiously affiliated facilities, making the laws vulnerable to freedom of religion claims. Holtzman, “How Crisis Pregnancy Centers Finally Met Their Match,” 94.
California has public programs that provide immediate free or low-cost access to comprehensive family planning services (including all FDA-approved methods of contraception), prenatal care, and abortion for eligible women. To determine whether you qualify, contact the county social services office at [insert telephone number].

The disclosure requirement for unlicensed facilities, provision #2, is a different notice altogether. To be in this category, the facility’s “primary purpose is providing pregnancy-related services” but “does not have a licensed medical provider on staff.” These centers also must offer two or more of the factors mentioned in the paragraph above pertaining to licensed centers, except they do not have to offer abortion services or contraception counseling. As a result, provision #2 applies only to CPCs and would not apply to comprehensive health clinics. CPCs in this category must post a status disclosure stating “[t]his facility is not licensed as a medical facility by the State of California and has no licensed medical provider who provides or directly supervises the provision of services.”

Legal scholars were hopeful that the FACT Act would be the disclosure ordinance that could finally stand up to constitutional scrutiny. In Have Crisis Pregnancy Centers Finally Met Their Match: California’s Reproductive FACT Act, written in 2016, prior to the Supreme Court’s decision, Holtzman predicted that “the California law will be held constitutional because it sufficiently overcomes the shortcomings of those invalidated or unimplemented ordinances.” It is important to understand why the FACT Act, in spite of its relative strengths, did not pass constitutional scrutiny. It strongly suggests that the possibility of regulating CPCs by mandating disclosures no longer exists, warranting examination of other regulatory strategies.

The Ninth Circuit ruled that CPCs engaged in professional speech, a category similar to commercial speech that is held to intermediate scrutiny. The Ninth Circuit defined professional speech as “speech that occurs between professionals and their clients in the context of their professional relationship,” which in this case, warranted intermediate scrutiny which must “directly advance a substantial governmental interest,” and the Ninth Circuit ruled that the state had a substantial interest in the health of its citizens. Holtzman concluded that “[i]f courts follow the Ninth Circuit’s holding and continue to classify the speech in question as ‘professional speech,’ then courts should find the FACT Act Licensed Notice provision meets intermediate scrutiny requirements and is therefore constitutional.”

**REPRODUCTIVE FACT ACT MEETS THE SUPREME COURT**

Holtzman’s 2016 predictions were not realized. The Supreme Court overturned the Ninth Circuit’s decision in 2018, invalidating the FACT Act. The Supreme Court argued that they had “never recognized professional speech as a separate category of speech subject to different rules.” While the Court has previously used lower scrutiny when evaluating a law that “requires professionals to disclose factual, noncontroversial information,” and also when “[s]tates regulate professional conduct that incidentally involves speech,” they argued that these precedents were not applicable to the FACT Act. According to the Court, the notice did not contain exclusively “factual and uncontroversial information,” since abortion was “hardly an ‘uncontroversial’ topic.” Consequently, the court applied strict scrutiny, which the Act did not pass.

The argument that the notice did not contain purely “factual and uncontroversial information” is one of many logical weaknesses in this decision. Contrary to the Court’s argument, the message that low-cost abortion and birth control exists in California is objectively factual. It is not a value judgement or a suggestion – it does not say “the state of California suggests that women use low-cost reproductive healthcare services,” which would be a controversial statement, because it is clear that not all parties agree that women should use these services. No one, however, can disagree that they exist.

The Court went on to propose that even if the Act had indeed only contained what it deemed factual and noncontroversial information, it still would not have survived intermediate scrutiny. Even though the Court acknowledged that the state did have a substantial interest in “providing low-income women with information about state-sponsored abortion,” it was “not sufficiently drawn” to promote this interest, as the notice didn’t apply to other, non-reproductive health clinics that service low-income women. Even though intermediate scrutiny requires that the law is “not more extensive than necessary to serve [the state’s] interest,” the Court took issue with the fact that the ordinance did not apply to health clinics that had nothing to do with reproductive healthcare.
more, the Court suggested that instead of “burdening a speaker with unwanted speech,” the state could inform women of its low-cost abortion options through a “public-information campaign,” ignoring studies that show that public health information campaigns tend to be less effective on low-income communities.\textsuperscript{125} The client could believe them and wait to schedule the abortion, perhaps saving money for the procedure. By the time they try to obtain an abortion, it is no longer legally available. They may attempt to self-induce the abortion, or receive an abortion from someone who is not a certified abortion provider.

This risk is not “purely hypothetical.” Studies have demonstrated that making abortion illegal doesn’t stop them; it merely makes them unsafe. According to a March 2018 Guttmacher study, abortion rates are similar between countries that heavily restrict abortion and countries with few to no restrictions. The abortion rate for countries that prohibit abortion is 37 per 1,000 women, and 34 per 1,000 women in countries without restrictions on abortions, suggesting that people find ways to get abortions even when it is not legal.\textsuperscript{129} Additionally, according to Guttmacher, “induced abortion is medically safe when WHO-recommended methods are used by trained persons, less safe when only one of those two criteria is met, and least safe when neither is met.”\textsuperscript{130} Furthermore, a March 2018 study published in the \textit{American Journal of Public Health} showed that women who were denied a wanted abortion were more likely than women who received an abortion to “experience economic hardship and insecurity lasting years,” which is reflective of Cherisse’s experience highlighted in the introduction of this paper.\textsuperscript{131} While it is true that more research should be conducted in order to gain insights about the extent to which the misleading tactics used by CPCs affect the decision making of pregnant people, the well-documented danger of minimizing access to abortion makes it clear that the risk is more than “hypothetical.”

\textbf{WHY THE SUPREME COURT IS UNLIKELY TO UPHOLD PRO-ABORTION POLICIES: BRIEF HISTORY OF ABORTION IN THE COURTS}

In spite of the FACT Act’s legal strengths, particularly those highlighted by Holtzman, it is fundamentally unsurprising that the Supreme Court ruled in favor of the anti-abortion group. NIFLA is one of a litany of cases decided in the past 40 years since Roe that have played a role in reducing access to comprehensive reproductive healthcare. It is situated within a larger trend of anti-abortion policy and rhetoric that, due to the current partisan nature of the Court, is unlikely to change in the near future. Though the Court is supposed to be an apolitical body, there is substantial evidence that partisan political will is often exerted through justices by the presidents who appointed

In 2012, more than 2.6 million California women were in need of publicly funded family planning services. More than 700,000 California women become pregnant every year and one-half of these pregnancies are unintended. In 2010, 64.3 percent of unplanned births in California were publicly funded. Yet, at the moment they learn that they are pregnant, thousands of women remain unaware of the public programs available to provide them with contraception, health education and counseling, family planning, prenatal care, abortion, or delivery. Because pregnancy decisions are time sensitive, and care early in pregnancy is important, California must supplement its own efforts to advise women of its reproductive health programs. In California, low-income women can receive immediate access to free or low-cost comprehensive family planning services and pregnancy-related care through the Medi-Cal and the Family PACT programs.\textsuperscript{128}

While there has not been extensive research into how many pregnant people delay or forgo comprehensive reproductive healthcare (abortion or otherwise) after visiting a CPC, the Court did not state that there needs to be significant documented evidence of risks. Rather, the risks simply need to be “potentially real.” The potential harms of the deceptive tactics used by CPCs are clear. For example, a CPC could engage in one of its documented delay tactics, perhaps purposefully misstating a client’s gestation date or lying about the gestation date to which abortion is legal in the state.
them, and abortion has become a litmus test of legal conservatism.\textsuperscript{132} By briefly exploring how courts treated issues regarding abortion in the years after Roe, it becomes clear that conservative-dominated courts acquiesce to partisan will and make decisions that are highly reflective of political ideology.

It has been extensively documented that the pro-life movement essentially began in response to Roe. Abortion was of little political consequence before \textit{Roe}; it was not a major issue in the 1968 presidential campaign. In terms of public opinion, Rosenberg notes in his book \textit{The Hollow Hope: Can Courts Bring About Social Change?} that “by the eve of the Court’s decisions in 1973, public opinion had dramatically shifted from opposition to abortion in most cases to substantial, if not majority, support.”\textsuperscript{133} It was only after \textit{Roe} that significant, well-organized opposition began. In her book \textit{Below the Radar: How Silence Can Save Civil Rights}, Alison Gash points out that “although grassroots efforts to relax abortion restrictions were gaining momentum across the states, an organized pro-life campaign had yet to form. \textit{Roe} is often cited at the catalyst for this movement.”\textsuperscript{134}

Conservative, anti-abortion mobilization was not only intense and sudden, but also highly effective in limiting access to abortion. One of the most consequential abortion decisions post-\textit{Roe} was the 1992 case \textit{Planned Parenthood of Southeastern Pennsylvania v. Casey}, when the Court decided a new standard on which state regulations could be decided: whether or not the restriction imposes an “undue burden” on a woman seeking an abortion.\textsuperscript{135} While the decision technically reaffirmed \textit{Roe}, Gash notes how the Court was able to follow \textit{stare decisis} and reaffirm \textit{Roe} while still setting the stage for restrictions. She notes that justices have ample room to “differentiate subsequent decisions, even if they are substantively similar, in ways that limit the court’s past rulings,” which the Court did by adding the “undue burden” rule.\textsuperscript{136} Gash writes that because of \textit{Casey}, “states now have far more latitude to regulate access to abortion services.”\textsuperscript{137} Indeed, after \textit{Casey}, there was a sharp uptick in the number of abortion restrictions enacted.\textsuperscript{138}

In their work on judicial policymaking, Banks and O’Brien discuss the theory of the “Republican ideological model” of judicial selection, where judges are appointed as symbolic instruments of presidential authority.\textsuperscript{139} That is to say, the most important thing in a president’s consideration of a potential justice is whether or not they share the president’s political preferences and policy agenda. Banks and O’Brien state that this model is “associated with President Nixon” and was “refined by President Reagan” and subsequent Republican presidents.\textsuperscript{140} Of the nine justices on the Supreme Court in 1993 when \textit{Casey} was decided, five justices were appointed by Nixon and Reagan, and two were appointed by George H.W. Bush.\textsuperscript{141} It is logical that when the Court opened the door to abortion restrictions in \textit{Casey}, it was not unrelated to the Republican partisanship of the presidents who composed the court; partisanship in which abortion was now a key issue.\textsuperscript{9}

Since \textit{Casey}, abortion has only become a more politically divisive issue and one of the key criteria on which presidents evaluate Supreme Court nominees. Trump has not been secretive about his philosophy on Supreme Court nominees. During his campaign in 2016, he made this statement regarding overturning \textit{Roe}: “[i]f we put another two or perhaps three justices on, that will happen. And that will happen automatically, in my opinion, because I am putting pro-life justices on the court.”\textsuperscript{142} This is precisely what has happened: since his election, Trump has added two highly conservative, demonstrably anti-abortion justices to the Court – Brett Kavanaugh and Neil Gorsuch. In February of 2019, both justices dissented in a Supreme Court case that blocked a Louisiana law that required abortion providers to have admitting privileges at hospitals.\textsuperscript{143} This precise type of restriction, referred to as a Targeted Restriction of Abortion Providers (TRAP laws), was ruled unconstitutional by the Supreme Court in 2016 in \textit{Whole Woman’s Health v. Hellerstedt}, as it presented the kind of undue burden to accessing abortion established in \textit{Casey}.\textsuperscript{144} The precedent was clear – laws requiring abortion providers to have hospital admitting privileges were unconstitutional, but Gorsuch and Kavanaugh chose to skirt the foundational principle of \textit{stare decisis} when it comes to abortion cases, sending a clear message about their hostility to abortion access.

**LITIGATION STRATEGIES**

**BATTERY AND INFORMED-CONSENT TORT CLAIMS**

In her 2018 article “Crisis at the Pregnancy Center: Regulating Pseudo-Clinics and Reclaiming Informed Consent,” Tenielle R. Brown presents a novel way...
to make claims against CPCs that has potential, even in light of the Court’s anti-abortion stance. Broadly, Brown argues that because clients at CPCs “submit to physical exams or ultrasounds under CPCs’ false pretenses,” these individuals could successfully pursue a tort claim of battery against the CPC. The argument is that while the client may verbally agree to touching or even an invasive vaginal ultrasound, this cannot constitute informed consent, as it was reached under the false premise that the CPC was a bona fide medical clinic. This examination will position Brown’s strategy into a broader framework of looking at social change through litigation, drawing particularly on Galanter’s 1974 article “Why the ‘Haves’ Have Come out Ahead: Speculations on the Limits of Legal Change,” and conclude by discussing why this strategy may work despite the Court’s resistance to pro-choice policy.

LEGAL CONTEXT OF BATTERY CLAIM
Brown argues that “[a]ny time a woman is touched by a CPC staff member, if the touching is only consented to through deception by the CPC, she should be able to prevail on a civil battery claim.” In order to prove a claim of battery, a plaintiff has to prove that it is more likely than not that three things occurred: (1) the defendant touched the plaintiff intentionally, (2) the touching was harmful or offensive, and (3) the touching not consensual. As mentioned before, CPCs employ a variety of tactics to create the image that they are a comprehensive health clinic, many of which include touching, some of which is deeply invasive. Brown argues that forms of touching, such as having one’s pulse taken, “but especially undergoing a vaginal exam or ultrasound,” could be grounds for a successful battery claim. While CPCs may not touch all of their clients—perhaps some leave after reading pro-life literature and sitting in the waiting room—Brown notes that it is activities that include touching, like having one’s pulse taken or undergoing an ultrasound, that are ultimately the most dangerous. These forms of touching cement the misconception that the CPC is a real medical center, making it more likely that the client will heed their medically inaccurate advice.

ELEMENTS OF BATTERY
An array of nonconsensual touching can be grounds

10 The touching also cannot be “privileged”, such as contact that takes place during a lawful arrest by a police officer. Brown, 249.

for a successful claim of battery. The touching does not have to physically injure the plaintiff if the touching was offensive. The tort law code states that a battery “grievance consists in the offense to the dignity involved in the unpermitted and intentional invasion of the inviolability of his person and not in any physical harm done to his body.” Though it must be proven more likely than not that the touching was intentional, the plaintiff does not have to prove that the defendant had malicious intent. Successful battery claims have been brought regarding “spitting in someone’s face, removing someone’s hat, and tackling someone too aggressively in a junior-high football league.” The standard for determining offensiveness is that the touching must be objectively offensive, and courts have been clear that context matters in determining offensiveness. “Customary [touching] in the course of life”—an accidental touch on crowded public transportation or a routine handshake done in the course of business dealings—would not be considered offensive. What happens in a CPC is clearly not “customary.” As Brown points out, “it is hard to imagine a touching that could be more offensive—asking a woman to expose her belly or submit to a vaginal exam.”

The crux of the battery claim here is that the touching was nonconsensual, because it was agreed to on false premises. The law has been clear on this: “if it is shown that the plaintiff reasonably misunderstood the purpose of the touching, due to misrepresentations by the defendant, then consenting to the medical exam or procedure will not bar her claim.” For example, it is well established that even if one consents to having their blood drawn under the impression that the blood would be used for medical purposes, if the blood is then used as evidence in a crime against them, the consent given for the blood draw becomes invalid. Brown puts it aptly:

Consent is contextual. A famous Torts treatise even uses a medical example to make this point. It states: the “plaintiff who consents to manipulation of her body in the belief that it is for medical purposes, when in fact it is only for the sexual gratification of the defendant,” can have a cause of action for battery. You could substitute “sexual gratification” for “attempting to do God’s work” and the same premise holds.

BATTERY CLAIM VERSUS MANDATED DISCLOSURES
One of the reasons that a claim of battery against
CPCs is so promising is that it “sidesteps the collective action problems and political willpower obstacles that have long hampered larger-scale attempts to regulate CPCs.” Apart from their failure in the courts, one major limitation of mandated disclosures is that it takes a political will to make them law—legislatures in conservative cities and states are extremely unlikely to pursue this.

Since courts have a documented history of hostility towards abortion, it may seem as though any attempt to use a litigation-based regulatory strategy is a long shot. However, a tort claim may prove to be uniquely positioned to sidestep some of the pitfalls that have plagued other abortion litigation. The way in which an issue or legal claim is framed can have significant impact on its successful implementation. Although they have been proposed in the scholarly context of regulating CPCs, battery claims are much less overtly about abortion access than an ordinance requiring that CPCs state that California offers low-cost abortion services. In her book *Below the Radar: How Silence Can Save Civil Rights*, Alison Gash argues that the degree to which advocacy is “visible” can seriously hinder its efficacy—high visibility can create debilitating backlash, while low visibility can circumvent that hostility. Framing laws that regulate CPCs in terms of abortion has made these efforts more susceptible to anti-choice hostility. Gash writes that “[a]dvocacy efforts that invoke well-established or broadly supported principles and symbols conjure histories, trigger emotions, and outline an array of policy alternatives for less-informed publics.” Abortion is a well-established political issue that triggers automatic divisiveness and backlash. By making attempts to regulate CPCs about abortion, like the FACT Act did, the partisan resistance is immediately prompted. Though the ultimate goal of a tort-based strategy would be to minimize the harm done by CPCs and improve access to abortion, the link between a battery claim and abortion access is much more distant. This makes it more possible that conservative courts would feel comfortable ruling against CPCs.

**BATTERY TORT STRATEGY & THEORIES OF SOCIAL CHANGE THROUGH LITIGATION**

Brown’s proposal of using the tort law claim of battery to regulate CPCs is promising, but warrants an in-depth analysis of the legal power dynamics that would exist if a private citizen were to bring such a claim against a CPC, which is likely a part of a larger, well-financed network of CPCs. As Galanter explains in his article, “Why the ‘Have’ Come out Ahead: Speculations on the Limits of Legal Change,” while the American judicial system is premised upon the idea that both parties start off on equal footing in a given legal proceeding, certain parties are able to slowly manipulate the law to be written in their favor. Galanter theorizes that those who use the legal system can be categorized into two general groups: one-shotters (OS), for whom legal action is rare and usually of great consequence, and repeat-players (RP), who are involved with similar types of legal action over time. The distinction between the two is more subtle and complex than merely those with resources and those without. Although OSs are more likely to have fewer resources than RPs do, the significant distinction is marked by the reason, method, and frequency in which these groups engage in legal proceedings. For example, following a car crash, the auto-injury plaintiff would be an OS, while the insurance company respondent would be an RP. The plaintiff is in court to pursue damages and will most likely not find themselves in litigation about another auto-injury claim again, unless they are in another auto accident. The insurance company, on the other hand, will be involved in many more cases involving automobile accidents. Importantly, Galanter acknowledges that the OS-RP relationship is on a continuum, and not a dichotomous pair.

The key distinction between OSs and RPs is that the RP plays the long game. Litigation may be built into their business models, and they have the resources to strategically select, win, and lose cases to ultimately shape the law to be on their side. This is possible because of the U.S. system of common law, wherein precedent rulings inform how subsequent judges inform their own decisions. OSs, however, are wholly concerned with the outcomes of their cases because these decisions likely have significant and immediate impacts on their lives. There is, in fact, a difference between what RPs and OSs consider favorable outcomes—the OS, unlike the RP, is “unconcerned with the outcome of similar litigation in the future.”

Where the insurance company may very well choose to settle a certain case if it would set unfavorable precedent to lose at trial, the auto-injury plaintiff may be seriously in debt from medical bills and take the settlement. As Galanter puts it, “the RP is a larger unit and the stakes in any given case are smaller (relative to total worth). OSs are usually smaller units and the
stares represented by the tangible outcome of the case may be high relative to the total worth.” Galanter notes that the OS is more likely to “adopt a minimax strategy,” minimizing the risk of maximum loss, whereas the RP can “adopt strategies calculated to maximize gain over a long series of cases, even where this involves the risk of maximum loss in some cases.”

Galanter’s theory of RPs and OSs is informative when considering the potential risks associated with Brown’s battery tort claim strategy. To be clear, CPCs are different from many of the RPs that Galanter describes, which include insurance companies, unions, automobile manufacturers, and landlords, among others. Galanter does not discuss whether private, religious organizations can be RPs, but CPCs demonstrate many of the qualities inherent to being an RP. While litigation may not be as clearly built into their business model as it is for an insurance company, CPC networks are well-financed and politically savvy organizations with significant legal support. NIFLA provides legal counsel to over 1,400 CPCs with the express purpose of working to “protect, equip, and defend pro-life pregnancy centers.” NIFLA claims that over the past 25 years, their “[l]egal guidelines, best practices, and procedures […] have proven effective in protecting member centers from legal pitfalls.” NIFLA exists because CPCs are litigation-prone and know that they will continue to be involved in litigation. CPCs are concerned with the longevity and durability of their movement, and thus would want to litigate strategically to shape the law to favor that durability.

The individual making a claim of battery against a CPC would be, in theory, an OS. Brown asserts that during visits to CPCs, individuals are harmed as the result of unwanted touching, resulting in pain and suffering that warrants damages. Theoretically, the plaintiff would have representative OS characteristics: they are concerned about receiving damages and are not planning to partake in this kind of litigation again in the future. If a CPC thinks that the plaintiff has a decent chance of winning the case and potentially opening up the CPC to a litany of lawsuits, there’s a good chance the CPC would offer to settle with the plaintiff. The settlement could be attractive to the plaintiff, who sued in the hopes of receiving monetary damages anyway.

Of course, the plaintiff that Brown’s argument is premised on would be interested in holding the CPC accountable for their actions in a way that goes beyond a settlement and would likely be interested in the outcome of the law. The complexity of finding the “ideal” plaintiff, however, warrants further discussion. The American Bar Association states that it is a lawyer’s fundamental “obligation zealously to protect and pursue a client’s legitimate interests, within the bounds of the law.” When a lawyer takes on a client with the interest of creating social change through litigation, they are essentially asking their OS client to adopt a RP attitude. As Katz writes in his 1971 article, “The Public’s Interest in the Ethics of the Public Interest Lawyer,” “[a]ll too frequently such lawyers regard their client as a technical necessity—one who must have before they can file a complaint—rather than a person whose welfare is at stake.” The tension between a “zealous” representation of the client’s interests and the ability to create social change through litigation arises, as Katz notes, when the attorney avoids settling the client’s claim out of court “in the hope of successfully litigating a test case.”

In order to remain within the ABA’s ethical standards, it would be necessary for the client to be genuinely interested in the broader social importance of their claim. This is absolutely possible, but there is a danger that the “client will be unduly influenced by his attorney. The very nature of the usual public interest client—one likely to be poor or uneducated—makes it probable that he is overly susceptible to his lawyer’s manipulation.” The lawyer may not intentionally attempt to convince the client to take a certain legal path; a variety of inadvertent actions could participate in pushing a client into a certain choice. Furthermore, claims like these can last months or even years. As Galanter notes, a common RP strategy is to drag out litigation until the other side runs out of resources or the process simply becomes too exhausting. Even if the attorneys are working pro bono, they are still expecting the client to be involved in the process for what could be a significant amount of time.

When considering potential pitfalls of Brown’s proposal, it cannot be overlooked that the pro-life movement has been documented to engage in intimidation and harassment against those who work to make abortion accessible. The National Abortion Federation reported an escalation of hostility and violence towards abortion providers in 2017; accounts of trespassing nearly tripled from 2016, death threats doubled, and incidents of obstruction (such as obstructing physical access to an abortion provider) rose from 580 in 2016 to over 1,700 in 2017. The Feminist Ma-
jority Foundation reported in 2015 an increase in the dissemination of “Old West-style ‘Wanted’ flyers” that feature the names and personal information of abortion providers. The posters are often titled with inflammatory messages like “Killers Among Us,” or “Wanted For Killing,” and they are often circulated and posted in the neighborhoods where the providers live. The FMF found that about 19 percent of clinics reported being targeted this way in 2010 and that number rose to 28 percent in 2014.

Figure 5: A “Wanted” style poster distributed by anti-choice activists. Dr. Gunn was murdered by an anti-choice extremist in 1993.

While those who work at abortion clinics are the biggest targets of harassment, anti-choice activists have also been documented to harass and intimidate patients who go into bona fide reproductive health clinics, like Planned Parenthood. A 2016 investigative report from Rewire showed how anti-choice groups are considering abortion by using a technology called “geofencing,” which is software that tracks a person’s smartphone and builds a database of places they’ve visited in order to create a profile of them. CPCs can buy access to this software in order to track who goes to Planned Parenthood. Anti-choicers have already used this information to intimidate; when Rolling Stone reporter Alex Morris visited a Planned Parenthood in Birmingham, she reported that “[i]n the few minutes I’ve been in the clinic, the protesters have figured out my name and address, and as soon as I come out the door, they start calling them out, loud and self-righteous, as they hold their camera phones up to my face.”

While the plaintiff in a battery suit would fall into neither of the aforementioned groups, it is clear that anti-choice groups are not afraid to play dirty. These groups have a clear and extensive history of harassing and intimidating abortion providers and those seeking abortion, to the point where real violence has resulted. The politics surrounding abortion clinics are different than those regarding CPCs, and there have been no documented cases of violence or harassment towards those specifically working to regulate or dismantle CPCs. However, the ‘pro-life’ movement is the overarching ideology behind both anti-abortion harassment and the proliferation of CPCs. A plaintiff in a battery case against a CPC, especially one who refuses to settle, would run the risk of becoming the target of these practices.

CPCS AFTER A SUCCESSFUL BATTERY CLAIM
A successful claim of battery against a CPC could be an effective first step in working to minimize the harmful effects of CPCs. After a successful battery claim, CPCs that continue to receive consent to touch clients under the false premise that they are a bona fide medical facility would open themselves up to lawsuits that would have the credibility of precedent backing them up. To avoid lawsuits, CPCs would effectively virtual “fence” in an area, like a store or park. When a smartphone goes into the fenced area with location services turned on, the geofence begins to track the phones unique ID number. Then, advertisers can target you with specific advertising they think you may be interested in – if you enter the geofence of a gym, you may see ads for nutritional supplements. Marketers then sell this information to interested parties – in this case, to CPCs. “Anti-Choice Tech Stalking: Activists Are Using GPS to Track Women Entering Planned Parenthood,” Salon, May 26, 2016, https://www.salon.com/2016/05/26/anti_choice_tech_stalking_activists_are_using_gps_to_track_women_entering_planned_parenthood/.
be forced to reveal to clients that they are not licensed medical providers, and that they are pro-life, religious organizations that do not condone abortion, which essentially accomplishes what advocates were trying to achieve in the mandated disclosure ordinances.

Initially, these reforms may only affect justice on a more individualized level—that is to say, only for those who successfully bring battery claims—but over time, repeated litigation can be important in creating more widespread legal change. The example of the evolution of sexual harassment law is a useful comparison. Similar to the novelty of using a battery tort to regulate CPC’s, sexual harassment law was born out of creative legal argumentation and interpretation. At the time when legal discourse surrounding sexual harassment began to emerge, there was no existing law stating that sexual harassment was illegal. Advocates began by arguing that sexual harassment was sex discrimination, the latter of which is illegal. The situation with a battery claim is parallel; CPC’s touching of clients is not illegal, but battery is. Early advocates against sexual harassment had to carve out a new space in sex discrimination law in order to make sexual harassment illegal, and they did this quite effectively. Pro-choice advocates could find similar success in arguing for a new interpretation of battery tort law.

STATE FUNDING AND THE ESTABLISHMENT CLAUSE

Brown’s argument of pursuing a tort law claim of battery against CPCs is novel and has promise. However, it is important to consider the limitations of this kind of strategy, such as the ethical drawbacks to finding a “test-case” plaintiff, and the possibly slow and piecemeal process of change. Pro-choice advocates should utilize a multi-pronged approach that relies on a variety of litigation and policymaking strategies in order to minimize the harmful effects of CPCs from all angles. While strategies like battery claims against CPCs work to prevent them from spreading misinformation, the most ideal strategy would be to shut them down altogether. For this reason, many scholars have analyzed the idea of challenging government funding of CPCs on the grounds that it constitutes establishment of religion, in violation of the First Amendment.

The Constitutional standard for determining whether a law constitutes establishment of religion is the Lemon test, established in the 1971 case Lemon v. Kurtzman, in which the Supreme Court ruled that state funding of private, religious schools violated the Establishment Clause. The Court created a three-pronged test that a law must pass in order to not violate the Establishment Clause: (1) the funding must have a secular purpose, (2) its “primary effect” may not inhibit or advance religion, and (3) it cannot “foster excessive government entanglement with religion.” The third prong has the most promise regarding funding of CPCs. Although many CPCs are overtly religious organizations that admit as much on their websites, and have been shown to pray with clients, many scholars have noted that it’s likely a court would still rule that the funding has secular purpose, since it was established in Planned Parenthood v. Casey that the government does have interest in fetal life.

Funding could pass the second prong of the test for similar reasons – despite the clear religiosity of CPCs, there could be a successful argument that “abortion alternatives” programs are family planning services and this is the “primary effect” that CPCs have.

The third prong is more promising due to the intertwined nature of the relationship between state agencies and religious CPCs. The Lemon Court noted that the relationship between the state and the parochial school teachers was problematic when it entailed “comprehensive, discriminating, and continuing state surveillance.” The problem was not, in fact, that the Court thought that the teachers would be “unsuccessful in their attempts to segregate their religious beliefs from their secular educational responsibilities,” but rather that the state’s involvement with ensuring that this separation was happening would constitute excessive entanglement.

The Court in Lemon was also concerned because a condition of the funding was that the government would inspect the school’s financial records in order to determine how much the school was spending on secular versus religious activity, stating that “[t]his kind of state inspection and evaluation of the religious content of a religious organization is fraught with the sort of entanglement that the Constitution forbids.” The Supreme Court has previously cautioned against direct funding of religious organizations in Walz v. Tax Commission: “Obviously a direct money subsidy would be a relationship pregnant with involvement and, as with most governmental grant programs, could encompass sustained and detailed administrative relationships for enforcement of statutory or administrative standards.”

Though the nature of the “administrative relationship” between the government and CPCs differs widely across the states, there are several alarming examples that illustrate the potential of excessive...
government entanglement. In Pennsylvania, for example, CPCs receive funding from the Department of Welfare (DPW), and the DPW monitors the CPCs annually to ensure they are in “compliance with DPW standards, and staffed with caring professionals and trained counselors.” This requirement parallels the one that the Court found unconstitutional in Lemon regarding government monitoring of both teaching and school expenditure reports. If a court were to view the DPW’s annual monitoring and oversight about staffing decisions to be a “sustained and detailed” relationship, then DPW’s funding of CPCs would violate the third prong of the Lemon test.

**DRAWBACKS OF USING THE ESTABLISHMENT CLAUSE**

Although the legal parallels between state funding of religious schools and state funding of CPCs are clear, there is reason to believe that the courts may nonetheless uphold government funding of CPCs if challenged. As McElroy states, “the Supreme Court has shown reluctance in recent cases to find violations of the First Amendment’s Religious Clauses.” This reluctance, combined with the court’s overarching hostility to abortion access, means it may be unlikely that this claim would hold up in court.

It is also difficult to assess how much damage it would cause CPCs to strip federal and state funding. According to the anti-choice website “Human Life International,” non-profit anti-choice groups raised about $1.06 billion in 2010, about 80 percent of which went to CPCs, meaning CPCs had access to about $848,000,000 in 2010. Stripping government funding of CPCs may force some CPCs to close their doors or reduce the services that CPCs are able to offer, which is certainly a step in the right direction to reduce their potential harm. However, the huge amount of private funding CPCs receive makes it clear that preventing government funding of CPCs would likely not lead to a widespread shutting of CPCs.

It is also necessary to explore the potential backlash that this strategy presents. Scholars have focused their attention on how backlash to court decisions, especially Supreme Court decisions, can significantly hinder the implementation of those decisions or future progress on the litigated issue. Some scholars argue that the backlash to landmark Supreme Court cases like Roe v. Wade and Brown v. Board of Education should caution advocates against going to the courts in the hopes of creating social change. South-ern opposition to Brown was so intense that, according to Gash, “in many ways, it had a lasting effect on the ability of Brown to meaningfully alter the landscape of education opportunity for black students.” As previously discussed, Alison Gash recommends in Below the Radar that advocates for marginalized groups use “below the radar” litigation strategies that, by avoiding visibility, minimize the possibility for damaging backlash. Making a claim that the Establishment Clause is being violated is not below the radar. This strategy combines two of the most controversial issues in the U.S. – religious freedom and abortion. A decision to ban federal CPC funding would likely garner immense media attention, giving CPCs a platform in which to ask for more donations. It’s possible that an effective post-decision PR strategy could end up bringing in more funding for CPCs than they received from the government in the first place.

**CONCLUSION**

This paper is centered on answering this question: when considering that mandated disclosure notices are no longer legal viable regulation options, what is the most effective litigation-based strategy that pro-choice advocates can use to minimize the ability of CPCs to promulgate misinformation? I argue that using a tort law-based litigation strategy is promising and avoids many of the significant pitfalls that have plagued other regulation strategies. It is important, however, to note that this paper has certain limitations. While I found strong data that the misleading and deceptive tactics of CPCs are widespread, most of that data relies upon investigations by state affiliates of NARAL Pro-Choice America, a highly partisan, pro-choice organization. However, these findings are confirmed by more independent investigations, one being a Congressional committee report conducted in 2006 and the other a research project by the Department of Obstetrics and Gynecology at the University of North Carolina at Chapel Hill. Another limitation is that while it is well documented that CPCs engage in deceptive tactics, there has not been significant research into how this deception affects those who visit CPCs. It is unclear...
how many pregnant people ultimately decide not to get abortions or delay abortions due to their visiting a CPC. However, major CPC networks claim to have “saved” hundreds of thousands of “unborn” lives in the past few decades, and that 80% of pregnant women decide not to get an abortion after visiting their facility. These claims are not only unverified and likely are grossly inflated, but also suggest that at least some pregnant people decide not to get an abortion after visiting a CPC.

Abortion is a fundamental right. Since the Court codified this right into law in Roe, anti-choice groups have worked tirelessly to make abortion less accessible, and they have been very successful. Pregnant people sometimes have to travel hundreds of miles in the snow and ice across state lines to receive an abortion because TRAP laws have forced the closer clinics to close. One of the most insidious strategies of the anti-choice movement is the proliferation of crisis pregnancy centers.

Part I explored the deceptive strategies of CPCs by using newly compiled data in order to strengthen the claim that CPCs engage in deception in order to attract clients to their clinic and then dissuade them from choosing abortion. I found strong evidence that out of 190 clinics across the country, more than half promulgated medical misinformation that linked abortion to breast cancer, infertility/injury, and mental health issues. Part I also explored the constitutional and free speech context of mandatory disclosure laws in five cities and counties across the country before examining NIFLA v. Becerra, the case in which the Court slammed the door to this regulatory strategy completely shut. The Court’s argument in NIFLA suffered from logical weaknesses, suggesting that partisan will was more of a driving force in the decision than stare decisis. There is strong evidence that partisan will is behind the Court’s history of hostility to abortion that dates back to the 1970s, when Roe participated in unleashing unprecedented backlash against abortion.

Part II of this paper explored the regulatory strategies that advocates should use in light of the failure of mandatory disclosure ordinances. Ultimately, the lower-visibility aspects of Brown’s tort strategy give it serious potential to create effective social change. Given its potential ethical implications, pro-choice advocates and attorneys should continue to analyze practices of using test-case plaintiffs in cases focused on social change to ensure client-centered interests. Advocates should also continue to investigate the possibility of ending government funding by arguing that the funding violates the third prong of the Lemon test – excessive entanglement between the government and a religious organization that is manifested by the administrative relationship that the government engages in with CPCs in certain states.

Crisis pregnancy centers are an affront not only to abortion access but also to the autonomy and independence of anyone who can become pregnant. These facilities are based on the idea that women are merely vessels for carrying fetuses and that those fetuses are more important than the physical autonomy of pregnant people. CPCs are insidious and understudied; even though these facilities have existed for decades, there is no data regarding the rate at which they dissuade pregnant people from accessing abortion. It is imperative that scholars continue to study the effects of crisis pregnancy centers and that the legal community keeps fighting their deception and lies.

Endnotes

1 Winter, “26 Women Share Their Abortion Stories.”
3 Ibid.
4 Ibid.
5 “Google Search,” Search terms: “pregnancy resources near me,” April 19, 2019, https://www.google.com/search?ei=MaPCXMeXBcSU0wLtoYPYBA&q=pregnancy+resources+near+me&oq=pregnancy+resources+near+me&gs_l=psy-ab.3..35i39l2.3447.3605..3836...0.0..0.176.339.0j2......0....1..gws-wiz.......0i71.23raVvpkTvQ.
7 “The Truth Revealed: North Carolina’s Cri-
37 Ibid.
39 Ibid.
40 Ibid.
41 “Google Search,” Search term: “pregnancy center,” March 20, 2019, https://www.google.com/search?q=pregnancy+center&btnK=Google+Search&oq=pregnancy+center&gs_l=psy-ab.3..35i39l2j0l8.635.2069...2161...0.0.0.307.3426.0j4j1i1....3.0...1l2j0i131j0i20i263.cunVKEOPBRA. https://www.google.com/search?hl=en&source=hp&ei=2bLCXPi6BfrC0PEPwpqQ8Ag&q=pregnancy+center&btnK=Google+Search&oq=pregnancy+center&gs_l=psy-ab.3..35i39l2j0l8.635.2069...2161...0.0.0.307.3426.0j4j1i1....3.0...1l2j0i131j0i20i263.cunVKEOPBRA.
43 “Los Angeles, CA | Your Abortion Choice,” accessed April 7, 2019, http://yourabortionchoice.org/los-angeles/?gclid=CjwKCAjwv6blBRBBeiwAihBM-YL4MxcN3thXMKBFgbdDq8zMkudt6V-VQaw7Ms70Ah2NBAq_iV83p-RoCOJYQAvD_BwE.
44 Ibid.
45 Ibid.
46 “Crisis Pregnancy Centers Lie: The Insidious Threat to Reproductive Freedom.”
48 “Learn – Birthright.”
50 “Crisis Pregnancy Centers Lie.”
51 Ibid.
52 Ibid.
53 Ibid.
55 “Crisis Pregnancy Centers Lie: The Insidious Threat to Reproductive Freedom.”
56 “NIFLA | National Institute of Family and Life Advocates.”
crisispregnancycenterreport.pdf.
61 Waxman.
62 Waxman.
64 Ibid.
65 Ibid.
66 Waxman.
68 “Crisis Pregnancy Centers Lie: The Insidious Threat to Reproductive Freedom.”
69 Waxman.
71 Ibid.
72 Ibid.
73 “The Truth Revealed: North Carolina’s Crisis Pregnancy Centers.”
74 Ibid; Bryant and Levi, “Abortion Misinformation from Crisis Pregnancy Centers in North Carolina.”
76 Richmond-Crum and Kleder, “Maryland Crisis Pregnancy Center Investigations: The Truth Revealed.”
77 Ibid.
79 “Crisis Pregnancy Centers Lie: The Insidious Threat to Reproductive Freedom.”
81 Ibid.
82 Ibid.
83 Ibid.
88 Ibid, 943.
89 Ibid.
90 Ibid.
92 Ibid.
96 Ibid.
97 Ibid.
98 Ibid.
99 Ibid.
100 Ibid.

102 N.Y.C. Admin Code. As cited by Holtzman p. 90.


107 Ibid.

108 Ibid.


110 Ibid.


113 Ibid.

114 Ibid.


117 Ibid.

118 Ibid, as cited by Holtzman, 105.


121 Ibid.


123 Ibid.

124 Ibid.


134 Alison Gash, Below the Radar: How Silence Can Save Civil Rights, 21.


136 Alison Gash, Below the Radar: How Silence Can Save Civil Rights, 21.

137 Ibid.


139 Christopher Banks and David O’Brien, Courts and Judicial Policymaking, 33.
140 Ibid, 132.
144 Whole Woman’s Health v. Hellerstedt, 579 U.S. (U.S. Supreme Court 2016).
146 Ibid, 248.
147 Brown, “Crisis at the Pregnancy Center,” 248.
150 Wagner v. Utah Dep’t of Human Servs., 122 P.3d 599 (Utah Supreme Court 2005). As cited by Brown, 250.
151 Brown, “Crisis at the Pregnancy Center,” 250.
152 Ibid, 251.
154 Brown, “Crisis at the Pregnancy Center,” 221.
155 Gash, Below the Radar: How Silence Can Save Civil Rights.
156 Ibid, 34.
157 Galanter, “Why the ‘Haves’ Come out Ahead.”
158 Ibid.
159 Ibid, 98.
160 Ibid, 100.
161 Ibid, 96.
162 Ibid, 100.
163 “NIFLA | National Institute of Family and Life Advocates.”
164 Ibid.
167 Ibid.
168 Ibid, 906.
169 Galanter, “Why the ‘Haves’ Come out Ahead.”
172 Ibid.
177 Ibid.
179 Lemon v. Kurtzman, 402 U.S.
180 Ibid.
181 Ibid.
184 Ibid, 464.
186 Gerald Rosenberg, The Hollow Hope: Can Courts Bring About Social Change?

Bibliography

Alcorn v. Mitchell, 63 Ill 553 (Supreme Court of Illinois 1872).
Austin LifeCare, Inc v. City of Austin, No. A-11-CA-875-LY (U.S. District Court for the Western District of Texas March 20, 2012).
Austin, Tex., City Code, 10–10 § (2012).

Central Hudson Gas & Elec. Corp. v. Public Service Comm’n, No. 79–565 (U.S. Supreme Court June 20, 1980).


Evergreen Ass’n, Inc. v. City of New York, 740 F. 3d 233 (Court of Appeals, 2nd Circuit 2012).


Koffman v. Garnett, 574 S.E.2d 258 (Supreme Court of Virginia 2003).


O’brien v. Synnott, 72 A. 3d 331 (Supreme Court 2013).


Seigel v. Long, 53 So. 753 (Supreme Court of Alabama 1910).


Wagner v. Utah Dep’t of Human Servs., 122 P.3d 599 (Utah Supreme Court 2005).


Chinese Urbanism as a Way of Life: Becoming Urban in Shenzhen’s Informal Settlement

Yunhan Wen, University of Chicago Class of 2019

Abstract

Contemporary urban studies in China tends to acquiesce to the rural-urban divide codified in the hukou system without critically reflecting on what “urban” means beyond a legal category. Inspired by the Chicago School, I call for a return to understanding “urbanism” as “a way of life” to investigate Chinese cities. Under this paradigm, the city is not defined by any legal or geographical boundaries but is a unique social milieu that urbanizes its residents into adopting certain ways of life. My ethnographic work in Baishizhou, an informal settlement in Shenzhen, uses parenting as a lens to unravel both Chinese urbanism itself and the mechanisms through which residents become urbanized. Sending their children to the same school as hukou-holding parents, migrant parents in Baishizhou are socialized via between-group contact into the same Bourdieusian field of parenting as their hukou-holding counterparts. In this parenting field, the suzhi narrative fundamentally configures how migrant families devise and justify their parenting strategies. Roughly translated as civility, suzhi is at the center of discourses that legitimizes China’s rural-urban inequality by equating “the rural” to “the uncivil”. While better-off migrant families can engage with symbolic struggles to construct themselves as citizens with good suzhi, families with little means can only succumb to the symbolic power of such narratives and become self-critical of their deficiency of suzhi. The contrasting responses attest to the same observation—that in the field of parenting, the desired goods for migrant parents are not just children’s success but also the attainment of an urban status, which is not so much associated with urban hukou as with a set of practices and subjectivities that culturally distinguishes itself from the stereotypical “rural.” Based on these observations, I conclude that (1) frequent between-group contact in the city is a primary mechanism through which urban residents are influenced by the social forces of the larger urban milieu; (2) instead of a legally codified category, “urban-ness” in China is a sociocultural variable that can be learned, performed, and contested.

Acknowledgments

I am forever indebted to my thesis advisor, Professor Marco Garrido, whose deep concern regarding the difficulty in theorizing cities in non-Western context inspired me to form this project in the first place. This thesis would not have been possible without his constructive feedback, his indefatigable mentoring, and his commitment to students’ projects, however underdeveloped they might be. I also want to express my apologetic gratitude to my preceptor Alex Brewer, whom I bombarded, often at ungodly hours, with millions of logistical questions and who, despite that, helped me locate sources of funding that was desperately needed for an international project. I want to especially thank my informants in Baishizhou, who accepted me as part of their life, generously shared with me their life stories and their everyday experience, and trusted me with their children.
China is experiencing the largest internal migration in human history. In the four decades since the 1978 market reform, millions of rural migrants have left their homes for better opportunities in the city. Although explosive urbanization as a result of massive rural-to-urban migration has been witnessed across the Global South since the 1970s, what makes China’s case so peculiar is the hukou system (also known as “the Household Registration system”) which was first established in 1958.1 Born against the backdrop of China’s Great Famine when innumerable peasants tried to take refuge in urban areas, the hukou system was, at the very beginning, designed to control population flows so that cities would not be overrun by rural migrants.2 Yet, to claim that the sole purpose of instituting the hukou system is to prevent rural migration is a hubristic statement and a half-truth. The hukou system, upon its establishment, immediately provides powerful and exhaustive civil registration that grants “eligibility” to 1.4 billion Chinese citizens for the state, citizens have been categorized according to their Place of Hukou Registration (PHR), which follows their parents’ registration and is thus usually their birthplace.1,3 Citizens can only access public education, medical care, and other public goods at their PHR, and freedom of movement is severely curbed as a result. Such legal differences immediately translate into socioeconomic differences, due to the vastly uneven distribution of resources between big cities and small towns and between cities in general and the broad expanse of the rural landscape. Although the hukou system has relaxed enough to allow rural-to-urban movement, neither the legal difference nor their socioeconomic consequences has been eliminated. Although migrants are technically allowed to stay in the city, they continue to lack access to public goods and their socioeconomic status, after generations of institutional inequality, poses additional challenges for their livelihood in the city.

The inequalities posited by the hukou system have long been the center of scholarly attention. While the understanding of such ramifications is of paramount importance for migrant studies in China, previous scholarship has overlooked migrants’ subjectivities and their interactions with the larger urban environment. Moreover, by acquiescing to the legalized urban-rural divide imposed by the party-state, this scholarship not only denies migrants’ already limited agency but also singles out Chinese cities as an outlier incompatible with theories prevalent in Western urban studies, an indefensible approach. I am advocating for the inclusion of Chinese cities (or cities in non-Western contexts in general) into the development of contemporary urban theories. Chinese cities should be seen by urban scholars as sites to test dominant urban theories for the betterment of theoretical development instead of as an outlier. To this end, we should first put Chinese cities in conversation with classic urban theories. This is the goal of my paper. Inspired by the Chicago School of Urban Sociology, I call for a return to understanding “urbanism as a way of life” and the city as a unique social milieu that urbanizes its residents into adopting certain ways of life.4 Although some specific theories (such as the famous/infamous “centric rings”) have become more or less obsolete, we should bear in mind that the fundamental concern of early urban sociologists from the Chicago School is to understand “the urban” as a unique social milieu that distinguishes itself from the “rural.” This understanding of urbanism makes migrant studies an especially fertile ground: by investigating how migrants shape their subjectivities—their consciousness, beliefs, and values—in response to institutional and discursive forces unique to Chinese cities, we can render legible Chinese urbanism as a way of life.

My contribution to this venture is a three-month ethnographic study of migrants living in Baishizhou, the largest urban village in Shenzhen, China. Urban villages refer to informal settlements in urban China that house millions of rural migrants. Marked by physical disorder and filthiness, urban

---

1 Under this logic, however, migrants’ children, who are born in the city, do not have the city as their PHR.

2 Although some
villages, alongside with their residents, have been stigmatized as the antithesis of modernity and civility and as the rural cancer in the urban scene. However, my ethnographic observation reveals that instead of being isolated from the larger urban environment, migrant parents in Baishizhou are competing in the same Bourdiesuan field of parenting as their hukou-holding counterparts. In fact, these frequent interactions with hukou-holders enter migrant parents into the field in the first place. By observing the symbolic struggles migrant parents engage with in the field and how they devise and justify specific ways of parenting, I aim to understand whether and how migrants’ subjectivities are shaped by forces in the urban milieu and, ultimately, what constitutes Chinese urbanism as a way of life.

LITERATURE REVIEW

CHINA’S GREAT MIGRATION AND THE HUKOU SYSTEM

The hukou system bifurcates the population to agricultural hukou holders (referred to as “rural hukou holders” below) and non-agricultural hukou holders (hereafter “urban hukou holders”). Residents in rural villages are registered as the former and designated by the state to perform agricultural activities, while urban residents—though similarly restricted to their PHR due to the location-sensitive provision of public goods—enjoy much more freedom of movement and upward mobility. Prior to the 1978 reform, the hukou system had stringently prohibited rural-to-urban migration—peasants who moved to cities without authorization would be driven back by force when failing to produce proper identity documents upon frequent police inspection on the street. Therefore, when the 1978 reform relaxed the hukou system to allow rural-to-urban migration, the flow of migrants into the city resembled a dam failure and the official documents started referring to this population as “the floating population.” Although the hukou system has relaxed enough today to allow movement, it continues to bifurcate and codify the population into the local and the non-local and create institutional barriers that prevent non-hukou holders from accessing local resources from medical care to education to the right to purchase houses in certain cities. However, despite the hukou hurdles, migrants from rural villages and smaller towns continue to move, as a result of the extremely uneven distribution of resources that concentrate in top cities such as Beijing, Shanghai, Guangzhou, and Shenzhen. Consequently, unlike residents in Rio’s favelas, Mumbai’s slums, and Tondo, Manila, who can vote and participate in local political activities, migrants in Chinese cities never settle down in the formal sense. Their existence is marked by legal transiency incarnated in the Temporary Residence Permit (zanzhu zheng) that every migrant is required to hold, which is a stopgap measure for civil registration that recognizes the existence of migrants without recognizing them as residents proper.

The existing literature on migrants is voluminous and provides a variety of angles and insights into the subject, including the destinations of migration, the characteristics of migrants from different areas, the impacts of migration on sending communities, and, the most well-covered of them all, various forms of inequality generated/reinforced by the hukou system. Among different topics and subjects, I will focus on migrant housing and their family life, especially education for their children who stay in the city. Such selection is informed by Fan and Chen’s (2013) study of the new generation of migrant workers born after 1980. Those younger migrants, who are now in their twenties and thirties, are better-educated than the last generation and are more likely to bring families to the city or get married in the city, in contrast to the last generation of migrants who migrated alone and whose ultimate goal is to bring money back to their hometowns. The presence of families suggests the beginning of a trajectory of becoming urban residents that aims at long-term stay; I expect that for a family with children, housing and educational opportunities for children are important factors in family decision—whether to stay, to return home, or to stay with children sent back home. I will then review the status quo on migrant housing and education resources for migrant children.

MIGRANT HOUSING, EDUCATION FOR MIGRANT CHILDREN, AND THE SUZHI NARRATIVE

Starting in the 1990s, migrant enclaves have emerged in the peri-urban areas in major cities, where local peasants rent vacant units in their self-built houses to incoming migrants. These suburban migrant enclaves are often referred to as chengzhongcun, or villages-in-the-city/urban villages and are regarded as uncivilized space marked by disorder, filthiness, and informality.

2 This distinction is made to put children in the city in contrast with the “left-behind children” who have been the focus of many scholarly effort. See Liang, “China’s Great Migration.”
ty.\textsuperscript{3,13} With a frenzied housing boom and an increasingly speculative real estate market, migrants are often denied access to formal housing and homeownership and almost exclusively live in urban villages or dormitories provided by factories in suburban areas.\textsuperscript{14} Scholars have tried to identify the ramifications of such spatial patterns. While network analysts suggest that migrants’ networks are still confined within their social activities and face restricted interactions with nonmigrants, others disagree and emphasize the concept of translocality and question the statement that migrant enclaves are isolated from the rest of the city.\textsuperscript{15} A banal answer would be “it varies from case to case,” but within the varied interactions migrants have with nonmigrants, the narrative of suzhi, roughly translated as “quality/civility,” often stands out.\textsuperscript{16} A common belief regarding sushi is that one should work hard to improve their suzhi and that only those with suzhi deserve respect and happiness.\textsuperscript{17} The narrative of suzhi proves to be a double-edged blade for migrants. On one hand, it justifies discrimination and stigmatization against migrants who normally have little formal education; on the other hand, it gives migrants a chance to appropriate the narrative as a counterstrategy toward discrimination.\textsuperscript{18} Nevertheless, stigma against migrant enclaves that frames the enclaves as filthy, disorderly, and full of good-for-nothings and the narrative of suzhi reinforce each other to justify migrants’ incomplete citizenship in the city. Housing thus remains a big hurdle for the “urbanization of rural migrants.”\textsuperscript{19}

To add insult to injury, education for one’s children is contingent on housing in urban China. There are several ways for children to go to school in Chinese cities. First, through the public-school system. Under China’s Nine-Year Compulsory Education System, every child should be eligible for nine years of primary and secondary education in their place of hukou registration; and every home address has a designated public school.\textsuperscript{20} However, this is not an option for migrants—first, public schools only ac-

\textsuperscript{3} “Suburb” is a loaded word that invokes the North American image of low-density, automobile-based, middle-class residential areas. Here I used “suburbs” to simply denote areas that are far away from urban centers but still within the administrative boundaries of the cities, and as Ren (2013) has described, Chinese suburbs is a kaleidoscope of varied landscapes with factories, luxurious villas, and migrant enclaves alike.
high school education, most rural youths are not eligible for the college entrance exam, although the exam provides a sure way for upward mobility.

THEORETICAL FRAMEWORK
A RETURN TO URBANISM—A NEW DIRECTION FOR MIGRANT STUDIES
My concern with the existing literature on migrants lies in the disproportional weight given to the *hukou* system. It takes courage to articulate this concern, since the *hukou* system is indeed the source of glaring inequalities and *hukou* status is almost a deterministic factor for one’s urban experience. With no intention to downplay the continued need to study *hukou*, I want to point out other possibilities for scholarly effort on migrants—in the words of Fulong Wu, “systematic research on migrants in the urban environment.”

Instead of stressing yet again the *hukou* hurdles and the transience and illegality thus brought to migrants’ lives, we need a well-delineated picture of the migrants’ everyday lifeworld that hopefully unveils the meanings given by migrants themselves to their urban experience.

Urban sociologists have long been preoccupied with the idea of “urbanism,” which Louis Wirth famously defined in 1938 as “a way of life” that distinguishes the city-dominated modern age from previous epochs. The specific urbanism that Wirth described—a way of life marked by impersonality and social isolation—has been criticized as totalizing and downplayed the continued need to study *hukou*, I want to point out other possibilities for scholarly effort on migrants—in the words of Fulong Wu, “systematic research on migrants in the urban environment.” Instead of stressing yet again the *hukou* hurdles and the transience and illegality thus brought to migrants’ lives, we need a well-delineated picture of the migrants’ everyday lifeworld that hopefully unveils the meanings given by migrants themselves to their urban experience.

Urban sociologists have long been preoccupied with the idea of “urbanism,” which Louis Wirth famously defined in 1938 as “a way of life” that distinguishes the city-dominated modern age from previous epochs. The specific urbanism that Wirth described—a way of life marked by impersonality and social isolation—has been criticized as totalizing and limited to U.S. cities in the 1930s. The mechanical nature of Wirth’s theory assumes automatic correlations between the kind of subjectivities formed in the urban environment and ecological variables such as number, density of settlement, and the degree of heterogeneity of the urban population. This totalizing view ignores the idiosyncrasy of individual communities that are the most common object of observation for contemporary urban sociology. In other words, critics of Wirth argue, with good reason, that the microcosm of each individual community can shape its residents’ subjectivities more decisively than the urban environment as a whole.

However, what is essential and remarkable about the idea of urbanism is that it recognizes the city as a *unique social milieu* that is different from a village or a hamlet. The task for urban scholars is thus not to dismiss urbanism as a whole but to identify what makes the city distinct—a fundamental inquiry questioning the ontological validity of urban sociology as a subfield. Claude Fischer brought individual communities back into the theory of urbanism by offering a subcultural revision to Wirth’s. Fischer argued that the city is a unique social milieu in that it constantly produces unconventionality, which is itself a result of the presence of numerous subcultures—a wide range of distinct social groups whose members share a defining trait. Both Fischer and Wirth regarded heterogeneity accompanied by large populations as a definitive feature of the city. For Wirth, such heterogeneity is the ultimate reason for impersonality and isolation, and for Fischer, it is generative; heterogeneity leads to between-group contact that, instead of breaking down communities, reinforces the boundaries of different social groups and intensifies their subcultures. However, at the same time as it can reinforce group boundaries, between-group contact could also lead to mutual influences that infuse one subculture to another. These two processes are inseparable from each other and their constant operation is what makes the city a unique social milieu.

Current studies on urban China can benefit from these perspectives, which pose a provocative question that should not be dismissed as “too Western” for scholars of urban China: despite the *hukou* system and state capitalism, what makes urban China a distinct social milieu so different from the rural area? In other words, what is Chinese urbanism as a way of life? This question puts urban studies in China in conversation with those in Western contexts because it brings the classical Western debate on the difference between the urban and the rural to the Chinese table—a debate that has been treated as a nonissue for scholars of urban China, because the difference is already codified in law by the *hukou* system. However, we should not be satisfied with this state-mandated bifurcation, which tacitly frames migrants as the rural outpost in the urban environment, when in reality, migrants are constantly interacting with the urban environment and are already a constitutive part of it. Therefore, we should recognize the essence of theories on urbanism—that is, we should see “urban-ness” not as a *hukou* category but as a *sociocultural variable*, shaped by the unique social milieu of the city that exerts social forces on rural migrants and urban residents alike. Under this paradigm, a new direction emerges for migrant studies that could aid our understanding of both migrants and urban China as a whole: we need to see how rural migrants, long perceived as the rural outpost in the urban environment who trespass the
urban-rural divide, shape their subjectivities in reaction to the social forces of the urban milieu. By understanding the formation and transformation of migrants’ subjectivities, we can identify the social forces that define Chinese urbanism.

Admittedly, my intention behind raising “Chinese urbanism as a way of life” as the fundamental framework is highly political. By recognizing “urban-ness” as a variable independent from hukou status, we acknowledge migrants’ legitimate existence in the city. In contrast, continued elaboration of various forms of inequality caused by the hukou system victimizes migrants without making a case for them to stay or giving them the agency to make the case themselves. This is especially important now when the suzhi narrative was used by the state to justify forceful eviction of rural migrants living in informal settlements: in November 2017, a huge fire broke out in Beijing in a suburban migrant enclave, whose density far exceeded the legal standard in Beijing’s building code;33 in its official response to the fire, the municipal government dismissed rural migrants as the didu renkou (low-end population) and immediately evicted hundreds of rural migrants without offering alternative housing options.34 However, the benefits of using urbanism as the framework are not limited to academic advocacy. Existing studies indeed touch upon how living in the city and interacting with divergent actors shape migrants’ subjectivities—their aspiration, their self-perception, and their values and beliefs. An articulation of how those phenomena are brought forth by Chinese urbanism will provide much consistency to what now feels scattershot.

Zhan’s ethnographic research in one of Beijing’s suburban migrant enclaves emphasizes the concept of chuang (roughly translated into “venture”), the rhetoric of which rural migrants use to connote the idea of breaking barriers to achieve otherwise unattainable goals.35 This rhetoric resonates with the suzhi narrative where one should not blame the society but one’s own lack of effort and it encourages rural migrants to face the barriers with various forms of entrepreneurship in the informal economy. Although it captures rural migrants’ agency in its epitome, Zhan also points out that rural migrants need to adjust their aspirations based on institutional constraints—in face of Beijing’s especially harsh hukou policy, rural migrants understand their stay as temporary and live under a tight budget to save money for a better life in the hometown. A similar picture of juxtaposing institutional barriers and cultural agency is depicted by Pei-chia Lan, who studied migrant families that enrolled their children in public schools after Shanghai announced in 2008 a policy that integrates migrant children into public schools.36 Trying to understand the level of integration of migrants, Lan distinguished cultural incorporation (the recognition of differences between host and migrants), from institutional incorporation (redistribution of rights and resources).4 While institutional incorporation is largely out of the hands of migrants, migrant children actively appropriate the rhetoric of suzhi to strive for cultural incorporation—migrant children intentionally behave better than their urban counterparts in class, perfect their mandarin, and even pick up the Shanghai dialect. Ling’s ethnographic project regarding migrant children in vocational high schools in Shanghai is a nice addition to Lan’s: while the education system reproduces social hierarchy by categorizing migrant children as “laborers”, they also incorporate migrant children into the formal system and provide opportunities for migrant children to gain urban habitus.37 Since vocational high schools are usually located in suburban areas where people of lower socioeconomic status concentrate, migrant children form close relationships with other marginalized urbanites who have low socioeconomic status despite having local hukou. Through such socialization, the boundaries between the rural and the urban become blurred.

In all three accounts, we observe the distinctive impact on migrants’ subjectivities brought forth by the urban milieu—I call this the urbanizing impact on subjectivities. To call this impact “urbanizing”, however, requires justification. In the three cases mentioned above, cultural agency of rural migrants unanimously reflect in the construction of an image that is civilized, heroic, and infused with suzhi to

4 I want to especially mention Lan’s differentiation between cultural incorporation and institutional incorporation as a potential alternative framework for this thesis that I ultimately decided to place under instead of on equal footing with the larger framework of Chinese urbanism as a way of life. Although social and cultural incorporations are highly useful analytical lenses that can offer a clear account of migrants’ subjectivities, it is less forceful when it comes to connecting studies on urban China with classic urban sociological theories in the hope of internationalizing the latter, a (political) project that I am trying to accomplish.
distinguish themselves from the stereotypical rural: for rural migrants in Beijing, such urbanized subjectivities manifest as the battle-tested vision of someone who has “ventured near and far,” who, even though they in the end return to their hometown, return with down payment as well as their “eyes…opened”; for migrant families in Shanghai, it is their internalized discrimination against their rural origin as something the children should stay away from, as well as the pride they take in their children’s capability of speaking proper mandarin; finally, for vocational high school students, it is their feeling of disgust, maladjustment, and abandonment when the vocational high school system, stigmatized as somewhere “only bad students go,” predestines a life prospect contradictory with their own understanding of the self.

Then, is the impact on subjectivities brought forth by the urban milieu ultimately a civilizing impact? This might well be the case after an exhaustive study of the role of cities in the rise of modern capitalism that this thesis does not task itself with. However, what I want to highlight is that association between urbanizing and civilizing, in the Chinese context at least, is itself the result of the discursive forces that symbolizes “the urban” as the crown of civilization. When migrants construct “civilized” images, it is a performance according to not a universal definition of civility, but a definition proposed by the urban milieu—a performance that, though by no means suggesting that civility cannot be authentic in rural migrants, reveals that civilizing is a façade for urbanizing. This distinction is important, because it is never this thesis’s intention to further the characterization of rural migrants as culturally deficient and uncivilized before they become “enlightened” by city.

My own contribution to this framework will be a case study in Baishizhou, the largest urban village in Shenzhen. While urban villages in Shenzhen have long been stigmatized as “pus-oozing cancer” and Baishizhou is particularly famous for being “a lair for criminals” and an uncivilized space isolated from the larger urban environment, my study of parenting in Baishizhou reveals that migrant parents participate in the same Bourdieusian field of parenting as middle-class parents with urban hukou. The competition between migrant parents and their middle-class counterparts sets off a process of “becoming urban” without obtaining urban hukou for migrant parents, who strive to prove their suzhi and civility that make them worthy despite their lack of urban hukou.

BAISHIZHOU, THE EXCEPTIONAL URBAN VILLAGE IN SHENZHEN, THE EXCEPTIONAL CITY
Promoted to municipal status in 1980, Shenzhen, then the proverbial small fishing village immediately north of Hong Kong, has become a top-tier megacity, a world-class financial center, and home to numerous internationally successful high-tech companies—all within thirty years. As the party-state’s earliest attempt at market economy, Shenzhen’s success informed urbanization and administrative policies in traditional megacities such as Beijing and Shanghai. Shenzhen is by all means an exceptional city. Unlike Shanghai or Beijing, which have traditionally remained major urban centers in the past century, Shenzhen is a city built from scratch, and consequently, preexisting rural villages are dispersed across the city—including in the urban core—instead of concentrated in the peripheral areas. In the process of urban expansion, the municipal government of Shenzhen had neither money nor experience to simultaneously expropriate all the rural lands. Instead, it chose to adopt a piecemeal way of development and circumvented preexisting rural villages, which, after a series of convoluted historical processes, became high-density migrant enclaves that form a striking visual juxtaposition with the modern high-rises surrounding them. Consequently, there is a distinction between urban villages in Shenzhen and urban villages in Shanghai and Beijing—if both are called chengzhongcun or village-in-the-city in Chinese, the denomination is more accurate in Shenzhen, where some of the biggest urban villages are located right in the heart of the city.

As a city built from scratch, Shenzhen has a population that is largely made of migrants and migrants-turned-hukou holders. As a result, Shenzhen’s hukou-related policies are exceptionally progressive among the top four cities in China. In Shenzhen, enrollment in public schools is decided by a point system where all school-age children should apply to the public schools designated to the place they live (a practice called “schooling nearby,” or 就近入学), as long as at least one of their parents (1) has Shenzhen hukou or Shenzhen’s Residence Permit; (2) has lived in Shenzhen for at least one year; (3) has paid social insurance (社保) fees for at least one year. Applicants will be ranked according to their “points” (积分) and admitted according to this order until the openings are filled: for example, children whose parents have purchased housing with full property rights will score
40 points, but children whose parents are renting can also score 15 points; children whose parents who have Shenzhen hukou can score another 20 points while parents with Residence Permits can have 15 points. The progressiveness of Shenzhen’s education policy is striking because it allows renters, who are most likely not Shenzhen hukou holders, to send their children to Shenzhen’s public schools, and it does not discriminate against renters in informal settlements, where titles are in the grey area.

My specific field site, Baishizhou, is the largest remaining urban village in the central area of Shenzhen, and the Urban Renewal process that aims to demolish and redevelop Baishizhou has already started, making it an ideal site—probably one of the last sites—to observe in close quarters the socio-spatial landscape of urban villages in Shenzhen. Since Baishizhou sits in the heart of the city and is surrounded by middle-class enclaves called “xiaogu” (小區), public schools assigned to Baishizhou are the same as the ones assigned to adjacent xiaogu. Although local hukou holders who are also homeowners in xiaogu have far more points, most families in Baishizhou can get their children into nearby public schools under Shenzhen’s “schooling nearby” policy. In addition, migrant children can participate in Shenzhen’s entrance exams and go to high schools in Shenzhen, unlike the situations in other major cities.5

However, progressiveness does not spell complete institutional incorporation. Although both can participate, hukou holders and non-hukou holders are treated differently in the high school entrance exam; for the same school, the score cutoff will be higher for non-hukou holders. Moreover, even in Shenzhen, the already exceptionally progressive city, Baishizhou is yet another exception. With the ongoing Urban Renewal campaign led by the municipal government, Baishizhou is one of the few urban villages left in the central areas and families in Baishizhou are socioeconomically better-off than families living in urban villages in suburban Shenzhen.6 In peripheral districts in Shenzhen with much more urban villages and thus higher need for public education, the capacity of designated public schools can hardly cover the entire population, although further studies are needed to articulate the difference. Nevertheless, the story of migrant families in Baishizhou, as I shall show, speaks to the possibility of becoming urban without urban hukou. Under Shenzhen’s progressive education policy that allows children of migrants and urbanites to receive public education on equal footing, migrant parents enter the same urban game of parenting as parents with urban hukou, where self-perception, class consciousness, and a mindset identifiable as urban are formed and contested.

THE URBAN GAME AND PARENTING AS A BOURDIEUSIAN FIELD

I used the language of “game” to allude to the field theory proposed by Pierre Bourdieu. Attacking positivism on the one front and the excessive cultural turn on the other, Bourdieu proposed field theory as a way to bridge structure and agency. A “field” is a socio-spatial arena where agents, with their distinctive social positions (hence distinctive habitus) and combinations of capital, struggle, maneuver, and compete according to the rule of the field for desired goods.46 But one cannot thoroughly explain Bourdieusian fields without first clarifying what Bourdieu means by capital and habitus. Bourdieu famously extended the notion of capital into three primary forms: economic, social, and cultural. The former two are straightforward: economic capital can be directly converted to money while social capital refers to social connections.47 Cultural capital is a more nuanced concept that exists in three states: the embodied state that refers to “long-lasting dispositions of the mind and body”, the objectified state that refers to the material possession of cultural goods, and lastly, the institutionalized state in the form of educational qualifications. Of interest here is the embodied state of cultural capital. Its meaning is close to that of “cultivation” and it requires constant work for self-improvement much like the process required to build muscles. The embodied state of cultural capital is a part of habitus, which refers to the set of ingrained dispositions that reflect the social position of an agent and inform their social actions, and cannot be transmitted instantly as economic capitals. Consequently, the possession of this state of cultural capital claims both “the prestige of innate property [and] the merits of acquisition” and is easily recognized as symbolic capital—that is, “unrecognized as capital and recognized as legitimate competence.”48 Symbolic power, the power to be recognized as natural and legitimate, is a key concept in Bourdieu’s theoretical edifice, since the struggle in

---

5 See “Migrant Housing and Education for Migrant Children” under “Literature Review”.

6 The Municipal Government of Shenzhen has been clearing urban villages for redevelopment in the past twenty years.
a field is ultimately, “symbolic struggles to appropriate distinctive signs…to conserve or subvert the principles of classification of [classified, classifying goods].”

“The cultural game” detailed in *Distinction* is a particularly interesting example of such symbolic struggles in a field, largely because culture—or others’ perception of one’s culture—is at stake and its fundamental social nature “simultaneously presupposes and demands that one take part in the game and be taken in by it.” In other words, participation in the cultural game is not a voluntary process marked by an explicit declaration upon entering the game—interest in culture instantly enters the agent into the game, who then “[joins] in the collectively belief in the value of the game” and lack of interest in culture immediately marks the agents as barbaric, a trait that no social groups concerning Baishizhou—they are similar to real estate consulting agencies, and nongovernmental agencies, news outlets, real estate companies, and different WeChat Official Accounts of local government agencies, newspapers (Shenzhen News, Southern Weekly, etc.) and different WeChat Official Accounts of local government agencies, news outlets, real estate companies, and nongovernmental groups concerning Baishizhou—they are similar to

**A CLASS MOBILITY STORY?**

A careful reader will find that I equate urbanites (or urban *hukou*-holders) with “middle-class residents of the city,” and this calls for explanation. Under the *Hukou* system, class status and *hukou* attainment are highly overlapped, especially so in Shenzhen: if possession of urban *hukou* in Beijing and Shanghai can be a result of ancestral residence unrelated to class status, possession of Shenzhen *hukou* means either affluence or college-level education. The overlap between class and *hukou* immediately begs questions: why then, is the parenting field an *urban* game instead of a *class* game, where lower-class migrants try to achieve middle-class status? This is by all means one of several interpretations, but I contend that there is more at stake than class status. For one thing, *hukou* is a much more salient status indicator than class for migrants in urban China. Migrant families that I spent time with usually refer to their urban counterparts people with *Shenhu* (abbreviation for “Shenzhen *hukou*”) instead of middle-class parents, while, interestingly, those parents with *Shenhu* use “middle class” to refer to themselves. The possession of urban *hukou* in Shenzhen thus holds symbolic value. This goes back to the rural-urban divide that frames the former as vulgar while the latter as civilized. Moreover, class mobility for the children is the end goal for parenting, yet our interest lies in the impact of the urban milieu on the parents, whose socioeconomic status is relatively fixed until the children grow up. The desired goods in the parenting field for migrant parents themselves is thus not ascension from lower class to middle class, but transition from being seen as rural people with low *suzhi* to being perceived as urbanites with civility.

**DATA METHODOLOGY**

My ethnographic observation during my two-month stay in Baishizhou, Shenzhen, constitutes most of the data used for this thesis, in addition to seven formal interviews. I also supplemented my ethnographic observations and interviews with information from local newspapers (Shenzhen News, Southern Weekly, etc.) and different WeChat Official Accounts of local government agencies, news outlets, real estate companies, real estate consulting agencies, and nongovernmental groups concerning Baishizhou—they are similar to
Facebook pages, but on WeChat, the single most widely-used messaging and social media app in China.

I stayed in Shenzhen from July 7, 2018 to September 8, 2018, and spent on average twelve hours a day in Baishizhou. During the day, I jotted down shorthand notes on a notebook and typed up the field note at night. It took me one week to find a single room available for rent in Baishizhou and another two weeks to finally move in; before that, I lived in Tianbei in Luohu District that is 50 minutes away from Baishizhou via public transit. I alternatively lived in Baishizhou and Tianbei during the stay. At the beginning of my fieldwork, I reached out via WeChat to Baishizhou Xiaoqu (Baishizhou Group), a group of advocates and scholars who organize cultural events and write online articles speaking in favor of Baishizhou and advocating for the residents’ right to the city. After attending one of their events organized for fifth and sixth graders in Baishizhou, I got in contact with some of the childrens’ mothers, whom I approached as a researcher and offered to tutor their children in English in exchange for spending time with the family and asking questions. I was then added to a community WeChat group, where one of the mothers recommended me to other parents. Thanks to her promotion, I was able to get in contact and spend time with seven families by the end of my fieldwork. From the end of July to the end of my stay, I spent on average four hours every day with different families and, while not teaching, I reached out to state agencies and indigenous villagers to schedule visits and interviews and participated in different community events organized by the state-funded community center in Baishizhou. The interactions I had with residents in Baishizhou were mostly preserved in the form of fieldnotes, except for one resident, whom I helped carry heavy boxes back to her home and who, upon knowing that I was a student researcher, agreed to be formally interviewed. The other six interviews I have include one of the five village heads in Baishizhou area, a block manager at the Community Workstation (a type of local-level government agency), a Baishizhou Group advocate, and three middle-class residents in a neighboring xiaoqu (gated enclave) who were all elected members of the Homeowners’ Association of the xiaoqu. For a better focus, I did not use my interviews with bureaucrats for this thesis.

Fieldnotes and interviews were transcribed and coded in Chinese to avoid the loss of meanings in translation. For my first cycle of coding, I open-coded the data and let potential categories and themes emerge. Since fieldwork and participation have already given me a sense of the repeated patterns, I created several nodes prior to my actual coding. As I coded in broad strokes information regarding families, I realized that “buxiban” or after-school tutoring programs is often mentioned both when families expressed financial stress and talked about education; moreover, “buxiban” often appeared when one family talks about other families. I then coded the material for a second time according to those more nuanced codes that revolve around “buxiban” and attitudes toward both one’s own parenting/educating skills and those of other parents. The software I used is Nvivo. Any quotes of the fieldnotes and interviews will be translated from Chinese to English by me, and all names are pseudonyms.

**BECOMING URBAN IN BAISHIZHOU, SHENZHEN**

**FAMILIES IN BAISHIZHOU**

As the largest urban village in Shenzhen, Baishizhou houses highly heterogeneous social groups who have different relationships with the urban environment and thus have different understandings of their “place” in the city. 7 It is home to low-income, ill-educated rural migrants who work as cleaners and petty merchants, but a considerable portion of its population are white-collar young professionals who work in high-rise office buildings during the day and return to their rental units at night. Neither group has a strong attachment to Baishizhou, for different reasons. For young professionals, living in Baishizhou is temporary; it is the first stop that provides a bed before better opportunities arise. In online chat groups, young professionals frequently post advertisements seeking to pass down the lease, since they have either found better housing options, or changed jobs and company locations. On the other end of the spectrum, for low-income, ill-educated rural migrants especially of the older generation, leaving might be the ultimate ending since it borders on impossible for them to put down roots and become part of the formal population. As a street cleaner in his fifties from Hubei Province told me, he and his wife (also a cleaner) are working in Shenzhen just to take advantage of their last few years of physical capabilit-

7 For more context of Baishizhou, please see “Baishizhou, the Exceptional Urban Village in Shenzhen, the Exceptional City” under “Theoretical Framework.”
ty, and they will go back to their hometown when their bodies can no longer take it.

Within this heterogeneous population, families with school-age children stand out as the one group that cannot be flexible about where they live. As I mentioned before, Baishizhou’s very location is the reason that migrant children can go to the same public school as children with Shenzhen hukou. However, such inflexibility should not be seen as a constraint, but rather a sign of long-term residency and stability aspired to by the new generation of migrants, who were born after 1980 and are more likely to get married in the city or bring children and families with them upon migrating. In fact, all the parents I spent time with belong to the new generation, but heterogeneity still marks this group.

I want to first describe five of the seven studied families by their different combinations of capitals (see Figure 1). Although social capital varies between different families, it is not a salient factor that shapes parenting style, and manifests itself more in the accumulation of economic capital (for example, kinship ties that make business easier). Therefore, I use economic capital and cultural capital to differentiate between different families. It should be noted that the assigned labels of “high,” “medium,” and “low” in Figure 1 are not informed by an absolute standard, but are only meant to offer a rough sketch of the relative amounts of capital they have compared with each other. Moreover, the assigned labels are not deterministic: even when both families have a low amount of a certain type of capital, the nature of the capital respectively possessed can be qualitatively different, and thus informs different actions and choices. Yet, without taking away the agency from individual families, I want to point out that similar combinations of capital are more likely to bring about similar urban experiences, and for this reason, I divide the families into two groups for detailed discussion: first, I compare the families of Lailai, Lili, and Jingjing to show active participation in the urban game allowed by relatively sufficient capital; second, I compare the Xiao Sisters’ and Wenwen’s families to show how insufficient amounts of capital, though denying active participation, allows passive observation of the game that nevertheless shapes parents’ subjectivities. I want to accomplish two goals with my analysis: (1) to find out the mechanisms through which rural migrants are influenced by the urban milieu; (2) to show that “urban-ness” should be regarded not as a legal category determined by hukou, but a sociocultural variable that can be learned and performed. In return, by looking at the content of this performance, we can discern what is perceived as “urban-ness” in the Chinese context.

<table>
<thead>
<tr>
<th>Family (by children's names)</th>
<th>Economic Capital</th>
<th>Cultural Capital</th>
<th>Level of Participation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lailai (three siblings; the second child)</td>
<td>High; Born in Hong Kong with citizenship</td>
<td>Low; Parents barely received formal education</td>
<td>In the Game</td>
</tr>
<tr>
<td>Lili (three siblings; the third child)</td>
<td>Medium; Family-run roadside liquor store</td>
<td>Low; Mother finished primary school; Father middle school</td>
<td>In the Game</td>
</tr>
<tr>
<td>Jingjing (The only child)</td>
<td>Medium-low; Father structural designer; Mother online seller</td>
<td>Medium-high; Mother graduated from vocational college</td>
<td>In the Game</td>
</tr>
<tr>
<td>The Xiao Sisters (Two children)</td>
<td>Low; Mother shop assistant</td>
<td>Low; Mother middle school; Father received vocational training</td>
<td>On the Bench</td>
</tr>
<tr>
<td>Wenwen (the only child)</td>
<td>Low; Mother shop assistant</td>
<td>Low; Mother middle school; Father middle school</td>
<td>On the Bench</td>
</tr>
</tbody>
</table>

Table 1: Summary of family attributes

**LAILAI, LILI, AND JINGJING: IN THE GAME**
Lailai’s, Lili’s, and Jingjing’s families form an interesting trio: Lailai’s family is the landlord of Lili’s family, and their respective roadside stores stand right next to each other; Jingjing goes to the same school as Lili and is classmates with Lili’s younger brother. For this reason, their mothers know each other quite well. By all means, Lailai’s family does not belong to the category of migrant families; all indigenous villagers are Shenzhen hukou holders, and their lucrative rental businesses enable their children to pursue better private education. In Lailai’s case, she was born in Hong Kong, attaining Hong Kong citizenship, and now attends a private school in Hong Kong. However, her family stays in Shenzhen as they regard Baishizhou as their home, and, as we shall see, their presence complicates the field of parenting for Lili’s and Jingjing’s parents.

Lili’s mother dropped out before she finished sixth grade and migrated to Shenzhen when she was sixteen, while her father only finished middle school.
When I wrote down my Chinese name for Lili’s mother, she could not recognize the characters, and when we communicated via WeChat, she preferred to leave voice messages to avoid typing and spelling. Neither parent speaks Mandarin well because of their heavy Chaoshan accent. However, their businesses both in Shenzhen and back in Chaoshan are doing well. The rent and utilities alone for their liquor store-apartment cost 10,000 yuan a month, but when I asked for a rough estimation of net incomes, they shrugged and said, “Well, people like us don’t do bookkeeping.”

Lili’s parents met in Baishizhou, but they got to know each other because they were both from Chaoshan; Chaoshan migrants have their own hometown-based network. In fact, the liquor store was passed down to them from a relative when they struggled with money.

Jingjing’s parents, on the other hand, have better cultural capital. Although both are from rural areas and have rural hukou, Jingjing’s father graduated from a vocational college (大专) and her mother came to Shenzhen with high school diploma and obtained an adult degree from Shenzhen University. Both parents speak good Mandarin, despite some accent. They have, however, less financial means—Jingjing’s father works as a structural designer, and her mother, who quit a well-paying clerical job due to health reasons, works at home as an online merchant on WeChat that earns on average 3,000 yuan a month. For the sake of Jingjing, they rented a relatively spacious 2BIL room that costs 3500 yuan a month. They do not have strong networks in Shenzhen. However, Jingjing’s mother takes pride in yige ren chuang (venturing alone) and repeatedly told me about the story of how she came to Shenzhen on her own with only 200 yuan when she was barely eighteen.

Both Lili and Jingjing go to the same public primary school designated to Baishizhou. Lili is a rising sixth grader while Jingjing was a fifth grader. Both children are soon to face admission into middle school. Although middle schools in Shenzhen still follow the policy of “schooling nearby” and accept renters and homeowners alike in their designated areas, some of the schools—especially the prestigious ones—will accept applications from non-designated areas. The school will then organize a special entrance exam for non-designated applicants, and those who pass will be enrolled in addition to students from the school’s designated areas. This practice has its ramifications: on the one hand, property value is speculatively high in residential areas paired with prestigious schools; on the other hand, it feeds the already existing hype of cram schools called buxiban (补习班) in Shenzhen. These special entrance exams, designed to select only the most promising students across the city, are impossible for students who have only studied the regular curriculum.

Shenzhen’s parents are well aware of the possibility of sending their children to a school better than the designated one, and it is the possibility alone that gives parents pressure and anxiety. The reason, as Mr. Yang (a homeowner in a xiaoqu that shares its designated public schools with Baishizhou) puts it, lies in the fundamental human desire for a better life:

People are always after a better life…there is this sense of competition when it comes to two things: your own career, and the education you can provide for your children…I’m not sending my children to [the school paired with Mr. Yang’s xiaoqu]…the source of students (生源) is too bad. I mean, I do not mean to discriminate, but some of the families who send their children there, have unsatisfying suzhi—which is understandable, but, um…[he was searching for words, clearly aware of what he was about to say might be controversial to me, so I decided to make things less awkward for him by saying “I understand, if there are better options, why not?”] Yeah, right.

In defense of urban parents, I found during my interviews that some indeed send their children to the school shared with migrant families, and in fact see...
this as an opportunity for their children to learn empathy and humility. However, they also acknowledge that almost all the middle-class parents they know, including themselves, sign their children up for different kinds of buxiban or other interest-based classes such as drawing, programming, and dancing—the latter because applications to prestigious schools are not all about grades, and cultivation of various interests has long been an indicator of talents and status. In this way, expensive buxiban is the metonymy for a grand vision of elite education, class distinction, and self-made social mobility and is normalized as a symbol for responsible parenting.

However, Shenzhen’s parents are not limited to middle-class homeowners who are also the urban hukou-holders—migrant parents from rural areas such as Lili’s and Jingjing’s are also agents in this giant field of parenting, and the sense of competition is only strengthened when their children go to the same school as students with urban hukou. Through their children’s education, migrant parents achieve between-group contact with their urban counterparts. From parent-teacher conferences to the quick exchange at the pick-up time, migrant parents and middle-class parents exchange information and disseminate values.

Jingjing’s mother is decidedly influenced by the buxiban vision. Jingjing’s mother lamented their inability to afford Buxiban based on her interactions with other families—migrant and middle-class alike:

I can’t send Jingjing to TAL or Sikaole (both are very huge buxiban brands) like others do…you know, one of her classmates—we call him “Little Sheep”—lives here in the village as well, but his mother—oh my god, I really admire that woman’s tenacity! Her husband is never around but she invests everything into her son—sends him to TAL in Futian District every day for buxiban. It’s a group of children from Jing and Little Sheep’s class. Several parents from Shijicun (the nearby middle-class xiaoqu) organized this group and they take turns to drive them every day from here to Futian District. One day they asked me, “Hey, Jingjing is a smart kid, what kind of classes is she attending? Does she want to join our group?” and the only thing I could do was to smile and find an excuse to leave.

Expensive cram school, homeownership in a xiaoqu, and a car to drive children around all symbolize a middle-class way of childrearing that is unmistakably urban, since their possession is often paired up with middle-class socioeconomic status.12 The embarrassment that Jingjing’s mother felt when she walked Jingjing, however, suggested that not only was middle-class parenting desirable, there was also a shame in failing to provide it. Unable to secure buxiban due to the lack of financial means, Jingjing’s parents did not stop at lamentation. Instead, they decided to take the matter upon themselves. During our first meeting, Jingjing’s mother—when describing what the couple had done for Jingjing’s sake—declared (not without pride), “I don’t believe I can’t teach my own child well! I want to see how this will turn out and who will do better at educating our children—me or those with money.” For her, the difference between herself and those with Shenhu lies merely in the money, not in suzhi.

Jingjing’s parents thus had to make sacrifices. The first thing they needed to sacrifice was their rural hukou in Sichuan and Hunan. According to them, this provided them with considerable dividends since they as rural villagers were shareholders in the collectively-owned village businesses. But in order to prevent Jingjing from facing score discrimination in zhongkao (the universal high school entrance exam), they decided to give up the rural hukou and apply for Shenzhen hukou. This is a lengthy but doable process for long-time residents holding college-level degrees. The second thing they sacrificed was their career. To give Jingjing additional lessons and education at home, Jingjing’s father requested to work at home. Likewise, after her surgery, Jingjing’s mother did not return to the company she used to work at, but instead chose to stay at home to work part-time in online sales. Failing to provide the ideal education represented by buxiban, Jingjing’s mother actively appropriated her cultural capital to compensate her financial incapability and tried to imitate the middle-class style of childrearing with her cultural capital. However, she is still very much enthralled by the symbolic power of buxiban, evidenced by the undercurrent of insecurities beneath her appearance of confidence and her pride in mothering. When I asked more about Little Sheep out of curiosity, Jingjing’s mother was alerted, “Why? Is his mother asking for your tutoring as well?” For context, I was somehow perceived as a really good English tutor because I am a student of a foreign institute. In

12 See “A Class Mobility Story?” under “Theoretical Framework.”
fact, what made Jingjing’s mother and other migrant parents contact me in the first place was the perception that I might be a substitute for regular buxiban that was worth trying.

However, what is at stake for Jingjing’s parents is not just class mobility for Jingjing facilitated by education, but also recognition of both their and Jingjing’s civility. The home lessons provided for Jingjing were not limited to extra classes in Mathematics and Chinese, but also a transmission of what Jingjing’s parents regarded as classical cultural goods to Jingjing:

Every week her father will bring her to Nanshan Library to borrow books and she finished them within several days. I only let her watch good TV programs that introduce Chinese poems and her father will play all those famous foreign movies for her. Cheap, silly entertainment such as Tiktok (a viral app) and melodramas (which are popular among students in primary school) is not welcomed in our household—Jing also just doesn’t like it.

This is one of the many conversations that we had about how Jingjing is more mature and has better tastes than her peers because of her immersion in classical cultural goods. This immersion is possible only because her parents are well-versed in those goods. Knowing that I was also tutoring Lili, Jingjing’s mother would often make references to her better financial situations while hinting that her child is better cultured:

I’ve heard that Lili’s mother always brings her to BBQ stands at night, right? Haha, we can’t do it. They also bought smartphones for all their four children. Our Jingjing only occasionally uses the phone her dad broke. But nowadays kids waste so much time on phone. All those dumb videos and TV shows…Lili and her siblings are always on the phone and it can really be detrimental…(starting to recount Lili’s younger brother’s low grades and naughtiness).

What Jingjing’s mother was trying to do is exemplar of what Bourdieu calls symbolic struggles for distinctive signs. On the one hand, she distinguishes herself from affluent urbanites by emphasizing her homeschooling that demands more cultural capital and parental sacrifices than simply sending children to expensive buxiban. On the other hand, she distinguishes herself from better-off migrants by demonstrating that both she herself and her child have better suzhi, are much more culturally sophisticated, and should not be categorized together with the rest of the migrant families.

If Jingjing’s family is special because of the unusually high education level of both parents, Lili’s mother, in contrast, has more or less the typical amount of cultural capital among rural migrants. She could not take teaching upon herself because of her lack of formal education and she is very aware of the disadvantage it might bring to her children compared to other children such as Jingjing. However, she justified her parenting by presenting her children’s accomplishments in worldly aspects.

I am uncultured (没文化), unlike Jingjing’s parents. I can’t teach my children as they do. But my kids are smart and quick-witted. They have been handling themselves pretty well. My oldest kid already starts to earn money and is about to take my second kid to Shanghai for the Disneyland!

Her oldest daughter goes to a vocational high school (which means low zhongkao scores) in Shenzhen, which, as Ling (2015) pointed out, is a stigmatized space reserved for “bad students.” However, Lili’s mother defied such stigmatization by pointing out the worldly capability of her oldest daughter. Moreover, though she does admit her lack of culture (文化), Lili’s mother denies its influence on her children by emphasizing their cosmopolitanism (traveling to Shanghai and visiting the Disneyland). On another occasion, she made a statement that differentiates the lack of culture and the lack of suzhi. Since their stores/homes directly face the major streets and are thus not mixed with the clusters of dense buildings inside the urban village, Lili’s mother does not see herself as a resident of urban villages proper. For instance, when I mentioned I was going to Jingjing’s place, she commented “Jingjing’s place? I rarely go to the inside. It’s too messy and dirty. You should be careful when going in.” In this way, Lili’s mother also engages with symbolic struggles to appropriate distinctive signs; by tapping into the narrative that characterizes urban villages and their residents as dangerous, Lili’s mother separates herself from the stigmatized population living in the stigmatized place.

However, she was not uninfluenced by the between-group contact between affluent urbanites and migrant parents. Because of the closeness of their stores and homes, Lili and Lailai are best friends. As
I mentioned before, although Lailai’s parents are not well-educated, they are “loaded with money,” in Lili’s mother’s words. Indeed, during my short stay, Lailai travelled both internationally and domestically and never hesitated to buy the more expensive snacks. Lili’s mother is very conscious of the difference between her household and Lailai’s and does not approve of Lili’s friendship with Lailai, thinking that Lailai is spoiled and Lailai’s family secretly despises theirs:

Lili just doesn’t get it. Sometimes they don’t want Lili to come upstairs to their living area. And Lailai is so spoiled and always asking Lili to run errands for her and do stuff for her. So inconsiderate!

She turned to ask me about what I thought of Lailai’s English and her ability to learn. I replied that English education was faster-paced in Hong Kong and Lailai knew more vocabulary; but when it came to new materials that neither child had learned before, Lili was the faster learner. She then smiled not without pride:

Not to brag about myself, but my children are all really smart and they react very quickly to new stuff. Well, but Lailai’s family got money and they could send her to Hong Kong and hire private tutors to watch her do homework every day. But I don’t have to do it. They (her children) can do it themselves… Lili is also my sweetest child. She is the one who saves candies for her little brother and for me. That’s why she’s so easily manipulated by Lailai!

(I then asked about buxiban since she mentioned tutors)

I have never registered them for buxiban before (she said this rather gingerly and looked at me in anticipation of scolding and criticism, and I responded by saying no happy childhoods should be full of buxibans to encourage her to talk) Right? So I only registered Jiajia (the second oldest child) for a short buxiban this summer since next year she will face zhongkao…and that’s important and I think my second daughter is the smartest among all my children. But I can’t afford additional buxiban for Lili.

Lili’s mother’s feeling of anxiety and pressure, and her concession of sending at least Jiajia to buxiban are the best pieces of evidence of her participation in the urban game that puts the urban status of participants at stake. The urban status is not about hukou but about a set of beliefs and dispositions that are usually associated with the civilized urban in contrast to the backward rural. This is exemplified in Jingjing’s mother where she engages with symbolic struggles to justify her inability to provide the elite urban education metonymized by buxiban. She dismisses not the institutionalized cultural capital (academic credentials), but the embodied state of cultural capital possessed by Lailai by pointing out her bad manners that speak to ill family education and thus bad disposition. By contrasting Lili’s sweet character and competent self-management with Lailai’s bad disposition despite expensive private tutorship, she hints at the success of her family education that, though unable to transmit culture (文化), can transmit good disposition.

By observing both Jingjing’s and Lili’s families, we can obtain a glimpse of the social forces brought to them by the larger urban environment. Although pursuing what is perceived as the best education for one’s children and justifying oneself in the case of failure is a universal phenomenon not limited to big cities, their specific perceptions on what counts as the best education and as effective justifications are influenced by social forces unique to the city. The first kind of such forces is the frequent between-group that socializes both migrant families into the urban game, who, once entered, join the collective belief in the buxiban-marked parenting style. The second kind is a discursive one that juxtaposes the civilized urban with the backward rural. When the urban game has implications about the nature of parents who fail, the stakes become especially high for both Jingjing’s and Lili’s mothers because of their rural background, but they manage to defy this hostile discourse by laying claims to respectable suzhi. While Jingjing’s mother can claim cultural sophistication, Lili’s mother praises her children for worldly competence and good manners that attest to successful transmission of good dispositions from parents to children. By doing so, they strive for an urban status—a status unrelated to hukou that allows them to be culturally recognized as civilized
THE XIAO SISTERS AND WENWEN: ON THE BENCH
Both Lili’s and Jingjing’s families have considerable cultural and economic capital. If the combination of capital they have allows them to actively participate in the urban game, then families with meager capital can only submit to the symbolic power of *buxiban* in the game and deem themselves as naturally inadequate and lacking in *suzhi*.

Yue, the mother of the Xiao sisters, is the only mother who cried in front of me during our very first meeting. Yue is not a single mother, but her husband works as a cook in Guangzhou and earns 4000 yuan a month. She wanted the family to reunite but the education policy in Guangzhou does not allow renters’ children to go to public schools and she had to stay in Shenzhen alone with her two daughters. However, she has neither money nor time for children’s education. She works as a shopping assistant in the mall next to Baishizhou and earns 3000 yuan a month—the rent alone is 2000 yuan. As a result, she could not think of luxuries such as *buxiban* and she felt increasingly left behind by her children. She did not understand their conversations or their entertainment anymore and she attributed this to her fundamental lack of culture:

> Whenever they told me that, ‘Mom, you don’t understand anything.’ I just feel punched in my chest. But what can I do? I am uncultured (*没文化*) and I am already exhausted by work [started sobbing]. That’s why I want to seek your help. I couldn’t afford other classes and teachers. But could you just teach them something, anything, about some basic methods of studying?

Neither of her children perform well at school, and her major concern lies in her daughters’ in-school grades. Passing the challenging special entrance exam was never on the agenda, and she only thinks of the designated middle school as her children’s destination.

A similar but more intense case is Wenwen, who had already failed his entrance exam. Wenwen was the only middle-schooler I tutored over the summer. In fact, he already graduated from middle school and finished *zhongkao*. His score was not high enough for him to go to public high schools, especially due to the higher score cutoff for non-*hukou* holders. Wenwen’s parents both live in Baishizhou, but they lived in separate apartments to separate Wenwen and his father because Wenwen’s father kept hitting him for poor school performance. Not surprisingly, when Wenwen grew older, beating became fighting. Neither parent has had formal education beyond middle school and they both work at the mall—his mother as a shopping assistant, and his father a lobby manager. When Wenwen’s mother approached me, she had just learned of Wenwen’s *zhongkao* grade. Denied acceptance to public high schools, Wenwen was left with two options: vocational high school, or private high schools for students who failed *zhongkao*—such private high schools are expensive and can only provide substandard education, but they are nevertheless high schools. Unlike Lili’s mother who was content about her daughter going to the vocational high school, Wenwen’s mother thought proper high school education was important. Yet, she did not know whether private high schools could give Wenwen such education. Moreover, she hesitated about paying for private high schools and lacking returns from this investment.

Both mothers frequently engaged in self-criticism. During the time between knowing Wenwen’s scores and the application deadline for vocational high schools, Wenwen’s mother repeatedly mentioned that she “understands nothing” and wanted me to point out a way for her son. Yue, on the other hand, constantly warned her children against “ending up like [her]”:

> You need to study hard! In ten years, even the bathroom cleaners will need to have diplomas and without studying you will end up like me. No money and no culture.

In contrast to better-off migrant parents who engage in counter-framing and symbolic struggles, Yue and Wenwen’s mother take the responsibility upon themselves and see “failure” of their children as a result of their failed parenting and their lack of culture and civility. However, although neither family actively contest in the urban game, they are not completely out of the ring. They are on the bench, watching the formal contestants and internalizing the standard with which formal contestants compete. With their limited capital, they could neither homeschool nor send children to *buxiban*, but they recognized the importance of cultivation of interests promoted by the urban environment and tried to imitate it in their own ways.

Yue signed her second daughter up for the school’s free martial arts club in hope of exposing her children to more interests and hobbies, and when the school was organizing a martial arts performance in
the adjacent city, Yue immediately bought the ticket. The trip was unfortunately cancelled, and Yue complained about having to return the tickets with only partial refunds. Her complaint suggests that buying train tickets (around 50 yuan) is not a trivial expense, yet she bought the ticket without hesitation.

For Wenwen, his mother tried her best to respond to Wenwen’s request of support for his newly-grown interests. Really interested in Anime and Manga, Wenwen asked for a tablet for digital painting. Without fully understanding what digital painting is and what equipment it requires, his mother nevertheless bought it for him—a shoddy, cheap product purchased from the street that never really functioned. The dysfunctional tablet for Wenwen is another piece of evidence that his parents are “so out of it,” but for his mother, this is an investment in hope of getting Wenwen interested in something “good for him” that he refused to take advantage of (she never fully realized that the tablet was dysfunctional).

Unlike children in better-off migrant families, children in the worse-off households played an important role in strengthening the sense of “sitting on the bench”—going to the same school as middle-class children and enjoying the modern amenities and entertainment in the city, migrant children in Baishizhou, via between-group contact, are socialized into having the “urban need” that often catches off guard migrant parents with less capital. When the school trip was cancelled but the tickets hadn’t been returned, the second daughter of Yue, naturally for a fourth-grader, suggested that “Well, then we can go travel like everybody else!”—an innocent request that made Yue fall into long silence. For Wenwen and his parents, one major source of conflict is the incompatibility between Wenwen’s self-conception as a “cool kid” with diverse interests and his parents’ failure to respond to it.

The story of migrant families in Baishizhou from Yue and Wenwen’s mother’s perspective reinforces my statement that the point of this thesis is not class mobility but how social forces unique to the urban environment impact its residents’ subjectivities. In a sense, Yue’s and Wenwen’s households are representative of migrant households that are considered isolated from the larger urban environment. However, in both cases, we see the urbanizing impact the city brings to their subjectivities. Both mothers approached me because they know the existence of buxiban and the benefits it will bring—a sign that they, like Lili’s and Jingjing’s mothers, have been socialized through between-group contact into the parenting field. Unlike the better-off migrant parents, Yue and Wenwen’s mother cannot engage in active construction of civility to strive for an urban status; on the contrary, their lack of capital makes them internalize the suzhi narrative and recognize themselves as inadequate, uncultured, and bad at parenting. However, such self-criticism is exactly the result of their participation in the urban game. This participation is further evidenced by their realization of the importance of cultivation of interests and their subsequent efforts to provide it within the limit of their meager capital. Instead of being isolated from the urban milieu, migrant families with little means are experiencing the social force unique to the city as much as better-off migrant households—even more so, considering the oppressive effects it takes on their self-perception. Though not contesting in the ring of the urban game, migrant families with lesser means are on the bench, observing the game, awkwardly mimicking the move, and criticizing themselves for failing to make it to the ring.

CONCLUSION: BECOMING URBAN WITHOUT HUKOU

My analysis set out to accomplish two goals. First, to show the specific mechanism through which migrants are impacted by the social forces in the urban milieu. Second, to show that “urban-ness” is a sociocultural variable that can be learned, performed, and contested.

For the first inquiry, I agree with Claude Fischler that frequent between-group contacts unique to the city, as a result of its heterogeneity, will have profound ramifications on how social groups define and redefine their boundaries. It is precisely through between-group contact that dominant urban culture is being disseminated and coercive social forces in the urban milieu is being exerted. In the case of Baishizhou, migrant families’ interactions with middle-class urbanites facilitates the former’s entry into the urban game, and through their symbolic struggles in the game and their gradual identification with the values and beliefs promoted by the field, migrant parents form and transform their subjectivities—their belief, self-perception, aspiration, and emotions—and thus urbanize themselves.

For the second inquiry, I found that suzhi fundamentally configures how migrant parents devise and justify their parenting styles. For better-off migrant families, suzhi is something that distinguishes them
from the stereotypical rural, and in order to achieve such distinction, they actively compete in the urban game and construct various—sometimes subversive—images of responsible parents with suzhi. For migrant families with little means, they internalize the suzhi narrative that equates the rural to the uncivil and feel defeated by their own deficiency of culture and civility, which dooms their capability of being good parents according to the standard metonymized by buxiban. The unspoken requirement for being urban in the Chinese context is thus not necessarily the possession of urban hukou, but the recognition of one’s possession of civility, cultural sophistication, and, in a word, suzhi. Suffice to say that the path toward this recognition is fraught with sociocultural contestations.

This thesis is but an initial step in a much larger intellectual project that aims to understand the multitudinous aspects of Chinese cities, and the limitation of my studies lies primarily in the limited scale of the project, the exceptional progressiveness of Shenzhen, and the unique position of Baishizhou within urban villages in Chinese cities. A similar study conducted in urban villages in the periphery of Shenzhen might yield very different findings. However, even though the specific urban game played by parents in Baishizhou is not a universal phenomenon, the emphasis on understanding urban-ness as a sociocultural variable stays and has broader implications. The most important message that this thesis wants to convey is that we will continue to benefit from thinking of Chinese cities not as outliers but as fertile grounds to test and develop classical western theories. In the example of the Chicago School, although the specific theories developed in the 1920s can very much be obsolete, we are reminded by Andrew Abbott in his response to Michael Dear’s invitation to a debate that the pioneers of the Chicago School, such as Burgess, Park, and Wirth, did not so much offer substantial theories as provide a set of fundamental questions that concern urban scholars till today—what is “the urban”? How are urban structures being defined by various ecologies and fields? Can we observe modernity and postmodernity through the lens of the city? Those questions are certainly relevant for Chinese cities, and they have not received proportional attention.

Endnotes

8. Liang, “China’s Great Migration.”
13. Wu, Zhang, and Webster, *Enclaves and Transient Urbanism*; Laurence J. C. Ma and Biao Xiang, “Native Place, Migration and the Emergence of Peas-

14 Youqin Huang and Si-ming Li, *Housing Inequality in Chinese Cities* (Routledge, 2014); Wu, Zhang, and Webster, *Enclaves and Transient Urbanism.*

15 Wu, Zhang, and Webster, *Enclaves and Transient Urbanism.*


18 Murphy, “Turning Peasants”; Lan, “Segmented Incorporation.”


20 Chen and Feng, “Access to Public Schools.”

21 Lan, “Segmented Incorporation.”

22 Ibid.


25 Ren, *Urban China.*

26 Yeung, “Higher Education Expansion.”

27 Wu, Zhang, and Webster, *Enclaves and Transient Urbanism,* emphasis in the original.

28 Wirth, “Urbanism as a Way of Life.”


31 Fischer, “Urbanism as a Way of Life: A Review and an Agenda”; Fischer, “Theories of Urbanism.”

32 Fischer, “Subcultural Theory of Urbanism.”


34 Ibid.

35 Zhan, “The Urbanisation of Rural Migrants.”

36 Lan, “Segmented Incorporation.”

37 Ling, “Bad Students.”

38 Zhan, “The Urbanisation of Rural Migrants,” 1533.

39 Lan, “Segmented Incorporation.”

40 Ling, “Bad Students.”


44 Ibid.


48 Ibid.

49 Bourdieu, *Distinction,* 249.

50 Ibid, 250.

51 Ibid.


53 Ling, “Bad Students.”

54 Andrew Abbott, “Los Angeles and the Chicago

Bibliography


Bourdieu, Pierre. “What Makes a Social Class? On The Theoretical and Practical Existence Of Groups.” Lecture delivered to the Dean’s Symposium on “Gender, Age, Ethnicity and Class: Analytical Constructs or Folk Categories?” at The University of Chicago, April 9-10, 1987; Translated from French by Loic J.D. Wacquant and David Young.


Huang, Youqin and Si-ming Li. Housing Inequality in Chinese Cities. Routledge, 2014.


Liang, Zai. “China’s Great Migration and the


THE EFFECT OF POWER ASYMMETRY ON REGIONAL COOPERATION: COMPARING ENVIRONMENTAL NEGOTIATIONS IN ASEAN, THE EU, AND MERCOSUR

ANN-KATHRIN MERZ, GEORGETOWN UNIVERSITY CLASS OF 2019

ABSTRACT
This comparative study explores the effect of power asymmetry on regional cooperation. I argue that this effect depends on three factors: the existence and corresponding intensity of a conflict of interest, the power resources of interested actors, and the ability of these actors to purposefully direct their power resources. This study explores these dynamics in three cases of regional environmental negotiations: the 2002 Haze Agreement in ASEAN, the 2020 Climate and Energy Package in the EU, and the 2001 Environmental Framework Agreement in Mercosur. While contextual variables greatly impact the nature of the negotiated agreement, I find that under certain conditions, power asymmetry supports a cooperative outcome by creating “drivers” and “influencers” that allow negotiations to overcome the initial conflict of interest. Nonetheless, given power’s contextual nature, the effect of the power asymmetry cannot be removed from the situation in which it arises.

ACKNOWLEDGMENTS
This study represents the end of a long journey that began in 2017, when I first drafted my proposal for the Circumnavigator Grant. Throughout the process of applying to the grant, preparing my research trip, conducting field work, and developing my thesis, I was astounded by the generosity and support I found in my community and around the world. When stretched past my limits, I benefited from the kindness of others. My deepest thanks go out to all you, for helping me learn and grow and for the magical experiences you gave me.

First, I want to thank the Circumnavigator Foundation and Georgetown University’s School of Foreign Service, especially Jim Whalen, Tracy Sancilio, Gregory Rider, and Dean Lisa Gordinier, for making this experience possible. I also want to honor the memory of Circumnavigator Leader Margaret Ellen Parke.

I am also grateful for my Georgetown community. Thank you to my thesis advisor, Professor Kathleen McNamara, who first inspired me to study regional cooperation. I leaned on her generous support in carrying out my field work, developing my theory, and overcoming the challenges of writing this paper. Thank you to the many professors, including Professors Jeffrey Anderson, Andrew Bennett, Matthew Carnes, Leslie Hinkson, Eric Langenbacher, Pamela Sodhy, Yuhki Tajima, and James Vreeland, who shared their advice, expertise, and networks with me.

Setting up interviews in nine countries was challenging. Again, I owe incredible thanks to countless people, including Mercedes Rivera-Wansart and Gerardo Wright in Argentina, Juan Bogliaccini in Uruguay, Diego Eugenio Pizetta in Brazil and Antoine Ripoll for Europe.

As I traveled, people shared their homes and their culture with me. I will not forget the kindness of Herlado DiBattista in Argentina, Candy Pravcek in Uruguay, Erin Kay, Marc De and his family in Germany, and Karissa Prayago and her family in Indonesia.
This study is about the effect of power asymmetry on regional cooperation. It asks questions such as: What factors define power asymmetry in regional cooperation? Is cooperation more likely among unequal powers? Who drives forward regional cooperation? Why do some regions cooperate and others do not? These questions have become increasingly important. Threats such as human trafficking, transnational crime, and environmental destruction cross borders, making inter-state cooperation necessary to address them.

Studying cooperation within regions also matters in light of changing global power dynamics. Power is shifting away from the United States to different points across the world, producing the slow emergence of regional power centers. Amidst these changing power dynamics, regions display varying levels of cooperation. This variation underscores that cooperation is not inevitable, but the product of a host of variables. This study focuses on one variable, power, and its effect on cooperation.

THE ARGUMENT
This study models the effect of power asymmetry on regional cooperation through three factors: the presence of a conflict of interest, the power resources available to actors invested in the outcome of the conflict, and the capacity of these actors to effectively implement their power resources. The presence of a conflict of interest makes power relevant by providing an incentive to use it. The significance of the interest at stake determines the resources an actor is willing to sacrifice. The availability of power resources combines with the ability to implement them to determine the likelihood an actor will succeed. Together these factors model the nature of a power asymmetry within the regional context, which helps explain the resulting cooperation or lack thereof.

Influential actors can support cooperation by mediating conflicting interests to drive a group toward compromise. When the power asymmetry favors actors interested in achieving a cooperative outcome, it supports regional cooperation. However, only a certain range of power asymmetry supports cooperation. If the power asymmetry grows excessive, the influential actor may pursue global power at the expense of regional cooperation or bully its neighbors into one-sided agreements that hurt long term relations.

While power undoubtedly shapes regional cooperation, absolute statements on its effect are limited by power’s multidimensional, contextual nature. The utility of a power resource depends on the issue, meaning power’s effect is context specific. This study therefore situates the power asymmetry within a regional context. This study compares environmental negotiations in the Association of Southeast Asian Nations (ASEAN), the European Union (EU), and the Common Market of the South (Mercosur). Power plays a role in each case, because the regional actors have different interests in an agreement. However, this study finds that of the three factors in this power asymmetry model (conflict of interest, power resources, capacity for effective implementation), different factors were decisive for each case study. The depth of the conflict of interest mattered most in Mercosur, as a lack of interest left the agreement without a champion. The material resources mattered most in the EU case, as Western EU states’ material wealth allowed them to carry the cost of an agreement. Finally, the ability to effectively mobilize power resources mattered most in ASEAN, where Singapore stood out for its ability to take advantage of the various forms of power held by different actors. Despite these differences, the case studies nonetheless highlight similar patterns that shed light on the different forms power takes in the regional context.

DEFINITIONS
This study distinguishes between harmony and cooperation. Harmony is defined by Robert Keohane as a situation where policies (driven by self-interest and without regard for others) automatically facilitate the attainment of others’ goals. In harmony, interests naturally align. Keohane contrasts this with cooperation,
which requires actors to “adjust their behavior to the actual or anticipated preferences of others,” through a process of negotiation. Cooperation is tied to conflict, as resolving clashing interests drives actors to use sticks and carrots to achieve their desired outcome.

Next, as power is an essentially contested term with a multidimensional nature (meaning that different issues are influenced by different power sources), no single definition captures all its forms. This study therefore uses multiple conceptions of power. The realist conception of power uses material resources to therefore uses multiple conceptions of power. The realist conception of power uses material resources to quantify economic and military power. Its simplicity makes it an appropriate starting point to create the first constellation of regional power relations.

To capture more nebulous expressions of power, this study employs Steven Lukes’ three views. All three views of power hold that for actors A and B, “A exercises power over B when A affects B in a manner contrary to B’s interests.” In the one-dimensional view of power, the conflict of interest is openly displayed through a moment of decision-making. In the two-dimensional view of power, the conflict of interest is hidden, as it refers to moments where “decisions are prevented.” The two-dimensional view recognizes that actors anticipate other actors’ responses in their decision-making calculus. Consequently, actors make decisions to avoid an observable moment of conflict, as when a powerful actor prevents a controversial issue from arising. Finally, the three-dimensional view of power, which Lukes endorses, highlights how A influences B “by shaping or determining his very wants.” The third view emphasizes the role of the “socially structured and culturally patterned behavior of groups and practices of institutions in excluding potential issues from the political process.” The third view of power in particular highlights the role of context in shaping an outcome. This study therefore uses negotiations framed by regional organizations to clarify contextual differences in principles, norms, rules, and decision-making procedures that may alter an outcome.

CASE SELECTION – WHY ASEAN, THE EU, AND MERCOSUR?

While the context of regional organizations serves the purpose of clarifying defining differences, the organizations chosen for comparison share two basic similarities. First, in order to be relevant to negotiations, countries needed to be prominent within their respective regions. In Latin America, where regional organizations have shorter institutional lifetimes, the question of relevance was particularly important. Here Mercosur stood out as one of the most advanced integration initiatives during the time of negotiations. Second, ASEAN, the EU, and Mercosur are all multidimensional organizations. Multidimensionality means the organizations developed a broad mandate to address issues from a variety of arenas, including those pertaining to economics, security, and the environment. Environmental issues were therefore relevant to the respective organizations.

CASE SELECTION 2 – ENVIRONMENTAL COOPERATION

Due to the multidimensional nature of these regional organizations, studying cooperation within them is too broad for the scope of this study. Thus, this study will focus on negotiations leading to specific environmental agreements: ASEAN’s 2002 Haze Agreement, the EU’s 2020 Climate and Energy Package, and Mercosur’s 2001 Environmental Framework Agreement. This study focuses on a different environmental issue in each region. While choosing the same issue would give the study a sense of coherence, this would have been misleading because the unique history and varying levels of economic development in each region give them distinct approaches to environmental issues. Although choosing a different issue does not solve for these differences, it nonetheless assures that at a minimum the negotiations are relevant to the region. Studying an environmental issue is also interesting because environmental cooperation was not the primary aim in the founding of any of these three institutions, making its emergence puzzling.

THE LIMITS OF STUDYING COOPERATION AND POWER

The challenging complexity of both power and regional cooperation inherently limits the certainty behind the conclusions of this study. For one, there are simply too many variables that affect the likelihood of achieving cooperative outcomes. It is neither realistic nor helpful to consider every single variable that plays a role within these case studies. While this study attempts to include the most important variables, the many variables not included leave room for error.

Similarly, power can be expressed in many ways. This is especially true due to the nature of the two-dimensional and three-dimensional views of power. The two-dimensional view acknowledges that pow-
erful actors may intentionally exclude certain issues from politics, while less powerful states may choose not to act on an issue in anticipation of a more powerful state’s response.\footnote{17} The three-dimensional view of power moves beyond the behavioral focuses of the first two views, and considers the way power embeds systems and institutions with biases that unconsciously affect the decisions actors make.\footnote{18} This means actors and systems exert power within the negotiations in invisible ways. This subtle aspect of power limits our ability to recognize all its forms.

THE VALUE OF THIS STUDY
Despite these challenges, studying power asymmetry within the regional context is important. While one intuitively expects power asymmetry to play a role in regional governance, it is scarcely covered in academic research.\footnote{19} This leads scholars to assume that the largest country in economic or military terms is dominant within the balance of power. But power’s multidimensional nature produces unexpected winners. Problematizing power asymmetry helps clarify who is the key agent or structure influencing an outcome.

Understanding power asymmetry is especially important for environmental threats, which unlike security threats, rarely spark the same urgency to act. This means action is often driven by a select group of actors. As shown by these case studies, the fact that smaller states can drive or inhibit progress on environmental issues is important for understanding action on these issues in the future.

Finally, this study adds value for its comparative nature. Comparison allows patterns to emerge from the clutter and noise of case-specific details. These patterns speak to the forces acting across regions, making the study not only relevant for its regional context, but for understanding forces across the world.

THEORY
A number of theories inform this study. To analyze cooperation, the role of governments is understood through the lens of liberal intergovernmentalism and bureaucratic politics. Next, the effects of power asymmetry are understood through arguments for and against hegemonic stability theory. Finally, the most important contextual factors are highlighted by collective action theory. These theories inform the model of the power asymmetry, which involves determining the existence and extent of a conflict of interest, the power resources available to relevant actors, and the actor’s capacity to mobilize these resources.

THEORIES OF COOPERATION
THE ROLE OF GOVERNMENTS IN COOPERATION
The role of governments is understood through Andrew Moravcik’s conception of liberal intergovernmentalism. Liberal intergovernmentalism assumes rational state behavior, a liberal theory of preference formation, and finally an intergovernmentalist approach to interstate negotiations.\footnote{20} The liberal theory of preference formation captures how domestic groups pressure a national government to respond to their preferences, influencing a country’s actions on the regional stage. The intergovernmentalist approach describes the bargaining within Mercosur, ASEAN, as well as within EU institutions like the European Council, because the nation state is at the center of this bargaining process. The intergovernmental nature of these organizations is prominently shown in important negotiations that occur among the heads of state of the member nations. As heads of state their focus lies in representing their people, meaning leaders arrive at the regional platform with national interests in mind.

The problem with liberal intergovernmentalism for power analysis, however, is the assumption of rational state behavior. While some countries develop a rational, coherent foreign policy in their political system, many others are too dysfunctional, corrupt, or polarized to achieve one. Because exercising power depends on a country’s ability to mobilize its resources, obstacles in the domestic political system can prevent the exercise of power and thus alter the expected power asymmetry. On the flip side, an exercise of power requires changing the target country’s decision-making calculus, meaning the rationality and cohesion of the target matters. In this process, the country exercising power has to make assumptions about what the target values to alter its decision. But this neat depiction of an exercise of power fails when the target is not a single rational actor, but a multiplicity of actors with divergent interests. As this domestic chaos often holds true in the real world, the rational state behavior model is misleading in analyzing power asymmetry.

The theory of bureaucratic politics better explains how a divided government’s foreign policy forms. Graham Allison and Morton Halperin explain that government action results from “bargaining among players positioned hierarchically in the government.”\footnote{21} Especially when government leaders or
ministries compete, a government’s decision is not a single rational choice, but the product of “pulling and hauling.”22 This domestic competition explains the seeming inconsistency in Indonesia’s foreign policy, where the ministry negotiating the accord does not have the support of the ministry responsible for implementing it. The theory of bureaucratic politics further explains how a country can appear powerful through its material resources, but due to domestic divisions fail to mobilize these resources effectively. The degree of unity a country can achieve in mobilizing its government behind a single goal further explains some of the surprising power dynamics in this study.

THE EFFECT OF POWER ASYMMETRY ON COOPERATION
Discussion about the effect of power asymmetry on regional cooperation is informed by hegemonic stability theory. Hegemonic stability theory casts powerful states as paymasters;23 the powerful actor or group has more resources (including coercive power) at its disposal to pursue an outcome beneficial to everyone irrespective of other participants’ contributions.24 In describing the formation of cooperative institutions, Douglas Lemke argues that these institutions are more likely to be created and maintained with the presence of a hegemon.25 A regional power is incentivized to support regional initiatives, because they are a means to enable national economies within the region to become globally competitive; this also strengthens the bargaining power of the region—and in turn the regional power—in multilateral organizations.26 The regional power thus has an interest in supporting the development of its region in order to extend its own global influence. Hegemonic stability theory explains why a power asymmetry through the rise of regional powers may support regional cooperation efforts.

Critiques of hegemonic stability theory argue that hegemony undermines cooperation. First, hegemony may lead powerful countries to believe that they can achieve their goals without the region.27 This takes away their incentive to set-up the rules and carry the costs of a regional institution. Power asymmetry may further undermine cooperation by making powerful countries seem threatening to their neighbors, causing smaller states to coalesce in opposition to them.28 Through the selected case studies, this study explores the possibility that whether or not hegemony divides regional powers from their neighbors is dependent on the level of power asymmetry. It is likely that a moderate power asymmetry supports regional cooperation, whereas extreme power asymmetry undermines it.

THE IMPORTANCE OF CONTEXT FOR COLLECTIVE ACTION
Collective action theory, as explained by Elinor Ostrom, highlights how different structural variables may affect the likelihood of collective action in each region.29 Among the many variables to consider, four are especially important for this study. First, a greater number of participants in negotiations increases the likelihood of a non-optimal outcome due to the decreased noticeability of a single input (which encourages free-riding) and higher transaction costs in reaching an agreement.30 Second, the extent to which individuals are interested in achieving a public good affects the likelihood of its provision. Mancur Olson argues that if one individual has a stronger interest in achieving a public good, perhaps because they receive a greater benefit, the likelihood of its provision increases. Third, when actions are repeated, information about past actions affects the strategy participants choose, causing behavioral patterns to develop.31 Finally, the ability to learn norms alters the nature of cooperation.32 For example, adherence to the norm of fairness increases the willingness of participants to contribute to collective action.33 While these four variables affect the likelihood of collective action, there are many others that this study does not include—one study identified thirty variables that affected the likelihood of collective action.34 This creates the potential for error. Nonetheless, as Ostrom puts it, “There is no way that one can analyze the entire ‘spaghetti plate’ of variables that have been identified and their interactions in a single empirical analysis.”35

EXPLORING POWER ASYMMETRY THROUGH THREE FACTORS
The aforementioned theoretical lenses are used to model the effect of power asymmetry on regional cooperation according to three factors: the existence and depth of a conflict of interest, the possession of relevant power resources by those interested actors, and the ability to mobilize those power resources in pursuit of that interest.

A CONFLICT OF INTERESTS: WHO WILL CHAMPION COOPERATION?
Power asymmetry first depends on the existence of a conflict of interest. A conflict of interest means that at least two countries have an interest at stake in a particular outcome, and that these two interests are
not compatible with each other. Resolving the conflict requires negotiation, during which actors may use their resources to influence one another. Here an actor’s commitment to the interest may be one of the sources of the power asymmetry by making a country willing to expend more of its power resources than the opposing actor(s). For example, an interest pertaining to state survival will motivate great sacrifice whereas a tangential economic interest will motivate only minor actions.

POSSESSING RELEVANT POWER RESOURCES: WHO HAS THE RESOURCES TO EXERT INFLUENCE?
Power asymmetry also depends upon whether actors have the resources to influence each other. Given that power is multidimensional, the nature of the issue determines the relevance of the power resource.\(^{36}\) According to Kenneth Waltz’s realist conception of power, the size of the population and territory, resource endowment, economic capabilities, military strength, and competence matter.\(^{37}\) Norms may also constrain the relevance of other power sources, especially those pertaining to acceptable state behavior for members of an institution.\(^{38}\) For example, military power is less relevant in the EU, because the EU is a “zone of peace,” which makes military action against member states unimaginable.\(^{39}\) Softer forms of power may also shape or determine another actor’s interests.\(^{40}\) Finally, the availability of power resources changes depending on alliances. All three case studies involve more than two actors, meaning that the parties in conflict can ally with each other to shift the balance of power in their favor. This may be necessary when small states want to overcome the usually dominant regional power.\(^{41}\) A variety of factors, including the extent to which an issue divides countries, a country’s reputation, and the strength of regional relations, will determine the extent to which a country more heavily invested in a certain outcome can make other countries align with it.

THE CAPACITY TO DIRECT POWER RESOURCES: WHO CAN USE THEIR RESOURCES MOST EFFECTIVELY?
Once the relevant actors and the resources available to them are established, the ability to effectively mobilize these resources affects the power asymmetry. While an actor may have a clear interest and access to an incredible amount of resources, the actor may not be capable of directing them purposefully. For one, a divided government may inhibit the attempt to direct these resources. As discussed above, bureaucratic politics casts light on the pulling and hauling involved in governmental decision making.\(^{42}\) The domestic conflict among ministries matters in part because the environment ministries have distinct priorities from the foreign affairs and economic ministries. Disagreement among ministries becomes inhibiting when it prevents a country from making or carrying out decisions. Conversely, a country whose government is clearly united behind a single goal will be more influential.

CASE STUDIES
CASE STRUCTURE
The following three case studies begin with context, then conduct the power analysis, and finally analyze the nature of power asymmetry and its effect on cooperation.

Because power relations depend on context, each case highlights the defining characteristics of the regional organization.\(^{43}\) This is necessary to recognize their unique norms and rules for agreement. Likewise, different levels of institutionalization affect whether the organization acts as an independent actor (European Commission), a mediator (ASEAN Secretariat), or merely a platform for governments to come together (Mercosur). Next, the case outlines important moments in the negotiations to show the path that led to an agreement. Because negotiations generally occurred behind closed doors, the case studies draw on any available data.

The next section entails the power analysis, which begins by analyzing the existence of a conflict of interest. Here the extent to which an issue matters is important. When a country’s interest is difficult to determine, its domestic policy is considered, because the extent to which domestic groups care about this interest will influence the power resources mobilized to support it. Next, the power analysis compares various power resources, beginning with economic and military power for relevant countries. Following William Wohlforth’s approach to economic power, I operationalize economic power as GDP and GDP per capita using the World Bank Development Indicators.\(^{44}\) Likewise, I operationalize military power using data from the Composite Index of National Capability (CINC) Score.\(^{45}\) The CINC Score aggregates military personnel, military expenditure, iron and steel possession, energy consumption, total population, and finally urban population into a single value to represent military power.\(^{46}\) I also separate out the components of the CINC score to analyze a particular variable. These
comparative graphs set the foundation for analyzing power dynamics within the region. Of course, as Luke’s three views of power demonstrate, actors use power in a variety of context specific, subtle, and even unconscious ways. While this makes it impossible to identify every use of power within the negotiations, I consider other forms of power like pressure from international organizations and NGOs, media attention and institutional support. Finally, I conclude the section by analyzing the effect of the agreement, as this often reveals more information about the nature of the power asymmetry and the actors’ interests.

Next, the case analyzes the resulting power asymmetry. This involves a qualitative judgement regarding which factors in the power analysis were decisive. This is also an opportune moment for describing some of the factors not considered or questions raised by the cases. Finally, the conclusion draws general conclusions about whether the power asymmetry supported cooperation. This process allows for the formation of general hypotheses regarding the power dynamics that support or undercut negotiations.

ASEAN CASE STUDY: THE 2002 HAZE AGREEMENT
This case explores the role of power in the negotiations of the 2002 Haze Agreement. The negotiations involved a conflict of interest between the country creating the haze (Indonesia) and the countries suffering from its harmful effects (especially Singapore and Malaysia). Singapore’s capacity to mobilize a variety of power sources turned the expected power asymmetry on its head: rather than a situation of Indonesian hegemony, favorable conditions allowed the combined power resources of the ASEAN member states, international organizations, and the media to overcome Indonesia’s resistance to a treaty. Indonesia consequently signed the 2002 Haze Agreement, where the costs of haze reductions fell on its shoulders. Yet, the treaty’s lack of tangible impact shows the limits of power, particularly when actors attempt to exert power on a government as corrupt and fragmented as Indonesia’s.

CONTEXT
ASEAN’S DEFINING FEATURES
ASEAN was founded in 1967 by Indonesia, Malaysia, Singapore, Thailand, and the Philippines to stabilize Southeast Asia, as the founding members feared internal instability would allow communist insurgents to rise. ASEAN enlarged between 1984 and 1999 with the new members Brunei, Laos, Myanmar, Cambodia, and Vietnam. Vietnam’s induction was celebrated as proof of ASEAN’s stabilizing effect on the region. While ASEAN began with a security focus, over time it became more multidimensional to address other issues. Although ASEAN first rejected environmental measures, which former Malaysian Prime Minister Mahathir bin Mohamad saw as “eco-imperialism” from the developed world, over time the need for environmental action resulted in shared biodiversity and haze initiatives in ASEAN.

ASEAN contrasts starkly with the EU and Mercosur due to the norms embedded in the “ASEAN Way.” The ASEAN Way consists of the principle of non-interference in another state’s domestic affairs, a commitment to cooperation through consensus building, and a preference for national implementation over a regional bureaucracy. In ASEAN, votes need unanimous support to pass. Unanimity reflects the desire to protect sovereignty, a reaction to the history of colonialism in Southeast Asia. Unanimity also limits the influence of powerful members, as they have to convince smaller ASEAN members to support their cause. Similarly, the preference for national implementation prevents economically powerful actors from covering the cost of an initiative. These norms slow down cooperation but serve member state interests by protecting their independence. With its focus on sovereignty and aversion to large bureaucracies, ASEAN has light institutionalization. An ASEAN official hence called the ASEAN Secretariat the “official note taker,” as it keeps track of past agreements and discussions. ASEAN is thus much more intergovernmental than the EU, in that it serves as a platform to bring the region’s nations together, while limiting the outgrowth of independent regional bureaucracies.

THE PATH TO THE 2002 HAZE AGREEMENT
The origins of the 2002 Haze Agreement lie in the 1997 haze crisis. While the region has a history of haze, the crisis in 1997 reached new proportions. Land and forest fires destroyed 11.7 million hectares of forest in Indonesia, twice as much as in 1994. This produced a massive haze cloud affecting 70 million people in Malaysia, Singapore, Brunei, Southern Thailand, and parts of the Philippines. The haze directly harmed the agricultural sector, which was responsible for roughly 20 percent of economic output in most Southeast Asian countries. The haze also disrupted air travel, drove health expenses, deterred tourists,
and more broadly harmed restaurants, hotels and other hospitality businesses. The negative economic effects of the haze were further exacerbated by the 1997 economic crisis.

The proportions of the 1997 haze crisis and the economic recession drove action. At the 1997 ASEAN Ministerial Meeting on the haze, ASEAN ministers produced the Regional Haze Action Plan (RHAP), an ad-hoc, non-binding accord to address the haze’s immediate impact. In the spirit of the ASEAN Way, the agreement asked member states to develop national plans to combat forest fires. Yet the countries suffering most from the haze, especially Malaysia and Singapore, wanted a comprehensive agreement to prevent future catastrophes. Malaysia and Singapore pressured Indonesia for more comprehensive measures, altering between harsh communication and milder strategies stressing the “spirit of ASEAN.” Similarly, at the 1998 ASEAN Ministerial Meeting, ASEAN ministers voiced their frustration over Indonesia’s irresponsible agricultural practices. Indonesia’s relations with Malaysia, Singapore, and Brunei consequently suffered.

In October 2000, ASEAN ministers began negotiating a new haze agreement. Five months later, the first intergovernmental negotiations consisting of ASEAN environment and legal ministers began in Malaysia. Negotiations continued the following months, culminating in an agreement jointly prepared by the ASEAN Secretariat and the UN Environmental Program (UNEP), which was presented at a conference in Malaysia. The agreement required participating countries to take legislative, administrative, and other relevant implementation actions. The development of a binding environmental agreement in a region characterized by an unwavering commitment to sovereignty speaks to the pragmatic limits of isolated national action on environmental issues.

POWER ANALYSIS
In order for power to be relevant to these negotiations, the ASEAN nations must have had a clear conflict of interest, access to power resources, and the capacity to mobilize them in the pursuit of their interests.

CONFLICT OF INTEREST
In the case of the 2002 Haze Agreement, Indonesia had a conflict of interest with the other ASEAN members. Indonesia had the clearest incentive to oppose an agreement, as it was the source of the haze and consequently would bear the majority of the implementation costs. Singapore and Malaysia were the primary actors in favor of regional haze cooperation, because their proximity to Indonesia combined with regional wind patterns made them the primary victims of the haze. The haze had severe consequences on human life and the economy for Singapore and Malaysia, creating desperate conditions that incited a government response. Singapore was especially invested in regional haze cooperation, as the small country could do nothing more domestically to combat the haze.

Brunei, and to a lesser extent Thailand and the Philippines, on occasion also suffered from the haze, as was the case in 1997. The haze thus created a conflict of interest between Indonesia, the source of the haze, and the other ASEAN nations, which suffered its effects.

POWER RESOURCES
The following section compares the power resources available to relevant actors, focusing on Indonesia, Singapore and Malaysia.

ECONOMIC POWER
The timing of the haze agreement after the 1997 economic crisis, as opposed to after earlier haze crises in the 1980s and 1990s, suggests that economic power played a role in negotiations. The graph below shows the extent to which the economic crisis wrought havoc on Indonesia’s economy, such that GDP even fell below that of Thailand. This economic weakness likely increased a sense of desperation in Indonesia that made it desire improved relations with its regional neighbors. This is reflected in comments by Indonesia’s former Foreign Minister Ali Alatas, who commented that the economic crisis made Indonesia aware of the extent to which the region is interconnected and interdependent. Wanting improved relations to manage economic issues would motivate Indonesia to resolve other sources of strife with its neighbors. Despite Indonesia’s temporary weakness, unity among ASEAN’s smaller nations was likely necessary to give them the combined strength to overcome Indonesia’s economic power. Especially considering ASEAN’s dedication to consensus, securing the support of the many smaller members would create a stronger platform in favor of an agreement. Even the small country of Brunei Darussalam, whose GDP barely inches above zero in the graph, was likely a valuable ally in the pursuit of unanimity.
Next, economic prosperity allows a country to exert influence through trade and investment. Singapore consciously uses its economic strength to raise its significance and in turn influence in the region. For one, trade data from 1998 to 2000 show that Singapore was the first or second most important trade partner for Cambodia, Indonesia, Malaysia, Myanmar and the Philippines.\(^6^9\) This makes Singapore’s success significant for these countries’ economies, giving Singapore a more robust platform to voice its concerns. Singapore’s dominance in trade is similarly reflected in its foreign investment, as Singapore pursues intraregional investment efforts. For example, in the 1998 Philippines-Singapore Action Plan (PSAP), Singapore promised aid in developing industrial parks, infrastructure, and tourism in the Philippines.\(^7^0\) This pattern continues to the present, as in 2018 Singapore was responsible for 69 percent of intraregional investment.\(^7^1\) Singapore’s advanced development makes it a valuable partner for fellow ASEAN members, which likely gives it influence across the region.

Economic wealth may also enable a country to drive negotiations by volunteering to host them. This was the case in ASEAN, as many meetings took place in Malaysia. Malaysia hosted the first meeting where the 1995 ASEAN Cooperation Plan on Trans-boundary Pollution was signed, the first haze negotiations in March 2001, as well as the meeting with the final signing of the 2002 Haze Agreement. Malaysia also had the advantage of presiding over the ASEAN Chairmanship in 1997, which made it the designated host and agenda setter for a variety of other ASEAN meetings.\(^7^2\)

While the aforementioned factors suggest the relevance of economic power for the haze negotiations, ASEAN’s structure is intended to inhibit this exercise of power. Recognizing that wealth can buy additional influence, ASEAN only permits countries to make equal monetary contributions to regional initiatives. The 2002 Haze Agreement hence establishes a fund to be administered by the Secretariat, to which each country contributes $50,000; the amount was later reduced to make it more accessible to smaller ASEAN members.\(^7^3\) The commitment to equal contributions limits the size of ASEAN initiatives as well as the influence a prosperous country like Singapore holds.

**MILITARY POWER**

Traditional indicators of military power, like the CINC Score, consider territorial and demographic factors, which cause the score to reflect Indonesia’s power. When considering military spending, which is also related to a country’s economic wellbeing, Indonesia’s expected advantage disappears. The graph below shows Indonesia’s military spending. The fluctuations in spending point to instability within the Indonesian government, while the dramatic drop beginning in 1997 speaks to the damage wrought by the economic crisis. This weakness may have made Indonesia vulnerable to pressure by its neighbors.

Indonesia’s weakness combined with Malaysia’s determination to address the haze was also shown by Malaysia’s 1997 decision to send 1,200 fire-fighters into Indonesia.\(^7^4\) This violated ASEAN’s norm of non-interference and embarrassed Indonesia internationally.\(^7^5\) The fact that Malaysia was not reprimanded for its actions speaks to Indonesia’s weakness. This event gave Indonesia a need to secure an agreement to safeguard its country from Malaysian deployment.\(^7^6\)
SOFT POWER: INTERNATIONAL ORGANIZATIONS AND THE MEDIA
Among power’s many forms, the soft power of international organizations and the media was also significant for the haze negotiations. The interest of international organizations was already shown by the Asian Development Bank’s early involvement in haze workshops and the UNEP’s help in drafting the final 2002 Haze Agreement. Singapore took advantage of the international environmental interest. To protect the relevance of the agreement, Singapore initiated the Sub-Regional Ministerial Steering Committee (MSC), which included international organizations, NGOs, and the private sector. Singapore again involved international organizations in 2007, when Singaporean Prime Minister Lee Hsien Loon brought the issue of the regional haze before the UNGA, an act which Indonesia criticized as domestic interference. Given that one of ASEAN’s goals is to isolate the region from foreign interference, Singapore’s choice to raise a regional issue on an international forum speaks to ASEAN’s limited effectiveness in the face of Indonesia’s power.

Singapore also employed the media and think tanks to put pressure on Indonesia. While Singapore’s government was restrained at first, during the peak haze crisis in September 1997, Singapore used a media campaign to pressure Indonesia. Similarly, various organizations within Singapore held dialogues and conferences discussing the haze issue. For instance, in June 1998 the Singaporean Environmental Council hosted a dialogue that brought together forty-five representatives from thirty-one relevant institutional and regional organizations, think tanks, academic institutions, NGOs, and private corporations. These efforts fit Singapore’s role in ASEAN as its “intellectual hub.” A Singaporean diplomat commented that Singapore takes prides in coming up with good ideas for ASEAN. By generating ideas on how to address regional issues, Singapore influences the shape of regional cooperation. This serves as a source of power as Singapore is able to generate discussion on issues that affect it, in this case the regional haze. This in turn raises the profile of these issues. Singapore’s actions demonstrate how power’s multidimensional nature benefits countries that can creatively activate various power sources, giving them more influence than expected.

THE TREATY’S EFFECT
The 2002 Haze Treaty did not change Indonesia’s behavior due to Indonesia’s internal corruption and fragmentation. First, corruption in the Ministry of Forestry delayed Indonesia’s treaty ratification until 2015, because the Ministry of Forestry sought to continue profiting from bribery and graft in illegal logging. This corruption was widespread, shown by a 2000 WALHI report that claimed corruption accounted for 80 percent of the forest concessions held by the country’s biggest timber tycoons. While the Ministry of the Environment negotiated and hence supported the treaty, the Ministry of Forestry saw a threat to its wealth and opposed it.

Implementation represented another challenge due to Indonesia’s fragmented bureaucracy. Nguitragool highlights that twenty agencies were responsible for controlling land pertaining to the haze. Furthermore, in January 2001 the Indonesian government enacted dramatic reforms that took power away from the central government and gave it to the provincial governments. Preventing haze consequently required action from many actors. To complicate matters further, Indonesia faces tremendous administrative costs as the government has to monitor its 6,000 inhabited Islands. This makes enforcement difficult, such that even when the Ministry of Forestry refused to renew forest licenses for firms, they simply continued operations. The combination of corruption and fragmentation undercut attempts implement the haze treaty, not to mention the possibility of succeeding in the effort.

THE RESULTING POWER ASYMMETRY
The negotiations of the 2002 Haze Agreement showed a clear conflict of interest between the actor producing the haze (Indonesia) and the actors suffering from its effects (especially Singapore and Malaysia). While Indonesia would at first seem like the obvious beneficiary of a power asymmetry, the vicious combination of the 1997 economic and haze crises unified the smaller ASEAN nations into a single front, making Indonesia vulnerable and empowering the smaller ASEAN members. Indonesia likely calculated that the combination of domestic weakness and external pressure made participation in a negotiated agreement preferable to the...
diplomatic costs of resistance. In this case, an opportune combination of factors allowed smaller nations to use the regional organization as a platform to pressure the regional power.

The extent to which the power asymmetry enabled Singapore and Malaysia was limited, however, by the difficulty of influencing the fragmented Indonesian government. A power asymmetry depends on a relative difference in power in a relationship between one actor and another. The exercise of power depends on the ability of the more powerful actor to influence the other. Indonesia’s fragmentation means it is not sufficient to influence the heads of government, but also requires influencing the implementing ministries such as the Ministry of Forestry. Especially in the absence of accountability due to corruption, this is increasingly difficult, especially for foreign actors. The case of Indonesia highlights the relevance of Allison and Halperin’s bureaucratic politics model for a power asymmetry. An actor, even when favored by a power asymmetry, will find it difficult to smoothen divergent interests when faced with a multitude of contradicting positions within another country. These factors, combined with the fact that Indonesia’s economic weakness from the 1997 economic crisis was temporary, limited the duration and the effect of the power asymmetry favoring smaller ASEAN nations.

Nonetheless, this case does not consider all possible relevant factors. For one, this case focused on the larger ASEAN nations. Yet ASEAN’s commitment to consensus means that even the smallest ASEAN nations are significant, highlighting how institutional structure can alter power dynamics.

CONCLUSION
The ASEAN case demonstrates that a power asymmetry is enabling when it creates privileged actors willing and able to drive forward negotiations. This role was taken on by Malaysia and Singapore. For both countries, the benefits of regional cooperation to reduce the haze outweighed the transaction costs of leading negotiations. This made them dedicated to mobilizing power resources. For one, their own economic power allowed them to host negotiations, as well as gave them significance as trade partners and investors that made the other ASEAN members more likely to support them. Second, Singapore was willing to involve other actors in the form of international organizations and the media in order to pressure Indonesia. Their mobilization efforts were invaluable, especially because ASEAN’s light institutionalization gives the Secretariat a record keeping role. When an actor is willing to invest extensive energy and resources into protecting a particular interest and circumstances allow them to do so, it drives forward regional cooperation.

The ASEAN case nonetheless underscores how contextual factors like regional norms limit the options available to actors. The combined values of non-interference and equality limited the reach of the 2002 Haze Agreement. The value of non-interference meant that Indonesia would have to address the haze, which as the country producing the haze, meant the cost of an agreement would fall primarily on its shoulders. With high costs and little rewards, Indonesia did not have much of an incentive to implement the agreement. Additionally, the value of equality meant that all ASEAN members should contribute the same amount of money to regional haze efforts, meaning the amount had to be reduced to accommodate the smallest members. This interpretation of equality, rooted in part by the need to safeguard each country’s independence, limited the impact of the agreement. As the ASEAN nations prefer sovereignty protection over rapid regional action, the agreement’s limited impact may have even been intentional. Nonetheless, the power of these norms show the additional challenges Singapore and Malaysia faced as they pushed for regional action.

EU CASE STUDY: THE 2020 CLIMATE AND ENERGY PACKAGE
The following case study explores the nature of power within the negotiations of the 2020 Climate and Energy Package. Negotiations involved a conflict of interest between the climate interest of the wealthier, older EU member states in Western Europe, and the economic interest of the less wealthy, new EU member states in Central and Eastern Europe. The power asymmetry favored the climate interest of the Western EU nations. First, the Western EU members’ economic power allowed them to shoulder the burden of emissions reductions as well as embed side payments in the agreement. Second, institutional support for the 2020 Package drove negotiations forward through a relentless search for compromise as well as gave climate policy normative power. Nonetheless, the need to accommodate the economic concerns of Central and Eastern European states also suggest that the power asymmetry was only moderate. This benefitted future cooperation by promoting satisfaction and compliance with the agreement.
The European Council Presidency likely benefited states’ strategic appointment of Commissioners. Like Mercosur and ASEAN, the EU began in part to pursue economic growth, but since then has become more multidimensional by pursuing a variety of other measures such as environmental policy.

The EU contrasts with Mercosur and ASEAN due to its heavy institutionalization. The most important institution for negotiating the 2020 Climate and Energy Package was the European Commission. The Commission is the EU’s main executive body. It institutionalizes member states’ interests through member states’ strategic appointment of Commissioners. Each Commissioner receives a particular portfolio, which includes the Directorate-General (DG) for Environment since 1981 and the DG for Climate Action since 2010. Because the Commission has the exclusive right of legislative initiative, it serves as a venue where member state interests compete early in the legislative process. As the Commission’s organizational mandate is the pursuit of regional cooperation, it relentlessly seeks compromise among competing interests that materializes in incremental agreements. These incremental agreements support the growth of EU climate policy.

The European Council was also significant, because it gave the 2020 Package high-level political support. The European Council consists of the heads of state or government of EU member states, the European Council President, and the President of the European Commission. Climate policy advocates held the European Council presidency throughout negotiations: Germany held the presidency in the first half of 2007, when the targets for the 2020 Package passed, while France held the presidency in the second half of 2008, when the legislative proposals completing the 2020 Package passed. Germany and France’s rotation in the European Council Presidency likely benefited the development of the 2020 Package.

The negotiations of the 2020 Package began with the growing public attention surrounding climate change in 2006. Public interest in climate change rose with the release of Al Gore’s documentary, “An Inconvenient Truth,” rising oil prices, and energy insecurity in Europe following the Ukraine-Russia energy dispute. Additionally, the British government released the “Stern Review,” a 700-page report concluding that unmitigated climate change could cost 5 percent of GDP annually.

Interviews with the European Commission suggest that this public attention created favorable conditions for the drafting of an ambitious climate policy. Meanwhile the Commission became more unified and committed to climate change policy. First, DG Environment and DG Energy advocated for a joint climate and energy policy. They framed discussion in an energy security narrative, which appeals to Central and Eastern European countries. The change initiative also benefited from new Commission leadership, because Catherine Day, the former Director-General of DG Environment, rose to become Secretary General; Day’s new position provided an invaluable connection to Commission President José Manuel Barroso. With this political momentum, the Commission drafted the Climate and Energy Package in the fall of 2006. It resolved a dispute between DG Environment and DG Energy versus DG Enterprise and DG Internal Markets regarding how ambitious climate targets should be, exemplifying the early compromise that made the Commission invaluable to negotiations.

Yet the proposed 2020 Package remained controversial due to disagreement among member states regarding how to distribute the effort, and hence the costs, of implementing the 2020 Package. Member states that had joined the EU in 2004 demanded increases to their allowed emissions to promote economic growth. These divisions almost led to the unraveling of the agreement. Nonetheless, the Commission found an acceptable compromise through the principle of fairness: member states with lower GDP per capita would be allowed to increase their emissions for economic development, while the wealthy member states would reduce their emissions. Although this approach is similar to the burden sharing from past agreements, the Commission altered the calculation method to ensure that twelve of the thirteen newest
EU members could in fact increase their emissions. This created a range where new members Bulgaria and Romania were permitted to increase emissions by 20 percent, while the richest EU members (Denmark, Ireland, and Luxembourg) would reduce their emissions by 20 percent from 2005 levels. With this compromise in hand, in March 2007 the European Council agreed to the 20-20-20 targets. In these targets the EU committed to a 20 percent reduction in greenhouse gases, a 20 percent increase in renewable energy usage, and a 20 percent improvement in energy efficiency by 2020. These targets created the framework for the legislative proposals of the 2020 Package.

Despite the acceptance of the March 2007 targets, new member states became more critical of the 2020 Package when it was revealed by the Commission in January 2008. To appease them, the final package passed in December 2008 included major concessions in increased emissions allowances for the energy sector as well as for the new post-communist countries, the latter representing a major victory for Hungary, the Czech Republic, and Poland.

POWER ANALYSIS
To evaluate the power asymmetry, the following section explores the divide in interests and resources between the wealthier, Western European nations and the newer, Central and Eastern European nations with regards to the 2020 Climate and Energy Package.

CONFLICT OF INTEREST
The developing EU climate policy became a contest between the environmental interests of wealthier Western EU members versus the economic interests of Central and Eastern EU members. Among Western EU members, the earliest support for climate change policy was demonstrated by the Netherlands, West Germany, Denmark, and Austria, when they unilaterally committed to reducing their emissions in 1989/1990 irrespective of other states’ actions. This began a pattern of differentiated commitments based largely on GDP per capita, which continued in the Kyoto Protocol and the 2020 Climate Change and Energy Package. The willingness of a few states to carry the cost of mitigation in the name of an EU agreement not only highlights their commitment to the environment, but also suggests a desire to use the EU as a platform for global influence. For these wealthy countries, a commitment to the environment meant that they would take action to address climate change regardless of other states’ choices. Acting as part of a greater whole, under the banner of the EU, also allowed these countries to claim an additional benefit in terms of global influence.

Resistance to climate policy primarily came from the ten new Eastern and Central European countries (led by Poland), who preferred to prioritize economic development. These countries joined the EU in 2004 and 2007; like then new EU members Spain, Portugal, and Greece in the 1990s, they argued that the new EU members should be allowed to increase emissions to catch up to their wealthier EU neighbors. These states did not believe tough emissions reductions were compatible with economic growth. Interviews with Romanian civil servants at the Ministry of the Environment revealed their concern with whether the country’s technological development was sufficient to allow for emissions reductions and economic growth. Similarly, comments from the former Prime Minister of Poland Donald Tusk showed a willingness to reduce emissions so long as they did not increase electricity prices and slow economic growth.

POWER RESOURCES
The following section compares the power resources available to relevant actors, following the West-East divide.

ECONOMIC POWER
Economic power gave certain states more legitimacy in negotiations. A former prime minister captured this difference in legitimacy in the comment: “Luxembourg can issue a veto once in a decade, but Britain once per week.” The graph below shows the GDP of a sample of six EU member states, illustrating the gap in economic wealth. The gap is especially dramatic considering the lines along which support and opposition for the 2020 Package were drawn. In this sample of six countries, Germany, France, Italy, and Denmark all supported an agreement. In contrast, Poland, the largest country in economic terms opposing the agreement, has little economic power in comparison to these countries.
Nonetheless, Poland played an important role as the leader of the opposition to the 2020 Package. Poland received this role in part due to the legitimacy economic resources gave the country’s leadership. The graph below shows Poland’s GDP in comparison to a sample of other central and eastern European states opposing the 2020 Package.

Poland’s distance from the other members, especially during the 2007-2009 window, suggests its natural leadership position. This hypothesis in part explains Poland’s authority to negotiate concessions with the French Council Presidency in 2008, when it secured the temporary derogation of auctioning in the energy sector.119

The potency of economic power is also shown in the differentiated targets and side payments in the 2020 Package. By using GDP per capita within calculations for emissions reductions, countries with higher GDP per capita likely had more say over parts of the package due to their role in shouldering the cost. While using a seemingly absolute criteria like GDP per capita may make economic wealth seem less relevant, the Commission’s ability to tinker with the calculation method to receive certain results shows the agency of the individual member states in choosing their emissions reductions. The 2020 Package also included side payments to less wealthy states by giving them more achievable renewable energy targets as well as auctioning revenues from the emissions allowance system as compensation.120 This allowed wealthy states to economically incentivize resistant states to acquiesce to the agreement.

On the other hand, the fact that the wealthy countries carried the burden of emissions reductions while the newer European members were permitted to increase their emissions also highlights the limits of their influence. Differentiated targets show a perceived need to accommodate the interests of the opposition; of the twenty-eight members’ emission targets, thirteen members were permitted to increase their emissions.121 The seemingly lenient margin for emissions reductions highlights the limits of economic power.

Military power

Military power is less relevant for the negotiations. The EU has established what Charles Kupchan calls “a zone of peace,” meaning the “level of interstate comity” “effectively eliminates the prospect of armed conflict.”122 Military power does not create threats among EU states, meaning it does not have the deterring effect found in more militaristic regions.

Like economic power, traditional metrics of military power, like population, affect a country’s legitimacy. When the Former French President Valéry Giscard d’Estaing spoke of French and German dominance in the European Council, he highlights the size of their population as one of their power sources.123 The graph below shows the populations of these six EU member states in 2007, demonstrating this gap.
The former Prime Minister of Luxembourg and former EU Commission President Jean Claude Juncker affirmed the importance of reputation: “Greater member states have a greater say. We never admit it of course, but one has to acknowledge that geography and demography are playing a role.” These statements show how metrics like population size matter in that they change member states’ perceptions of each other.

Military power could become more significant should climate change discussion increasingly follow security narratives, as the Commission first did when it linked climate change to energy security. In this scenario, France becomes more important. The graph below shows France’s military spending. It outranks Germany, as Germany’s history has made the country more reluctant to develop its military power.

Figure 6: Military spending by EU member states, 1990-2010

OTHER FORMS OF POWER — THE SUCCESS AND FAILURE OF NORMS
Given the many institutions and actors within the EU, the 2020 Climate and Energy Package was affected by a variety of other power sources. Among them, norms played the greatest role through the principle of fairness and the importance of EU climate leadership.

First, the normative power standing behind the principle of fairness was essential to the success of the 2020 Package. Fairness addresses concern among the opposition, shown by the Visegrád quartet’s (the Czech Republic, Hungary, Slovakia, and Poland) September 2008 joint statement emphasizing the need for climate policy to be constructed fairly. Additionally, as collective action theory posits, when participants think a proposal’s costs and benefits are fair, they become more willing to contribute. Fairness likely made wealthy EU nations more willing to shoulder the burden of the agreement. Allocating emissions on the basis of fairness is also clever, as the criteria of GDP per capita was on a growth trajectory in the EU. This means countries that are permitted to increase their emissions on the basis of GDP per capita can expect pressure to reduce their emissions once their GDP per capita rises. For example, Spain was permitted to increase its emissions in a 1990 burden sharing scheme, but experienced dramatic GDP per capita growth such that it committed to reducing emissions ten percent in the 2020 Package. Given that the Commission tracks precedents as the basis for future cooperation, these countries will be held accountable to their “fair share” of reductions in the future.

Next, normative pressure also infused negotiations in the form of a commitment to global climate leadership. The narrative of the need for EU climate leadership developed primarily after the United States withdrew from the Kyoto Protocol, as the EU needed to rally other industrialized countries to save the agreement. As this narrative’s origin precedes the accession of the Central and Eastern EU states in 2004 and 2007, it was likely more difficult to weaken this self-imposed commitment. Not surprisingly, Bürgin concludes that the pursuit of climate leadership served as a source of normative pressure over the opponents of the 2020 Package. The narrative of global climate leadership is especially effective, because it allows member states to frame the passage of the proposal in terms of a greater good that extends beyond the EU itself.

THE PACKAGE’S EFFECT
The Commission monitors progress on EU-wide agreements and is thus able to address shortcomings. This institutionalization reinforces EU agreements and gives member states an additional incentive to comply. At the end of the Barroso II Commission, it seemed as though the EU would reach the targets it set for 2020. This speaks to the success of the many incentives embedded in the package, as well as the value of having wealthier member states carry costs.

THE RESULTING POWER ASYMMETRY
The negotiations of the EU’s 2020 Package demonstrated a conflict of interest between the environmentally conscious Western EU nations and the economic interests of the Central and Eastern EU nations. The greater economic power of Western EU members legitimized their influence over the process, suggesting the power asymmetry pointed in their favor. Moreover,
their wealth enabled them to shoulder the primary cost of the initiative, which speaks to the benefits of their wealth. But this also raises questions as to what extent they chose to accommodate the Central and Eastern EU countries, or were forced to because of limits of their negotiating power. The power asymmetry thus favored the Western EU nations, but not so much that they were able to ignore the interests of other EU member states, forcing negotiation and compromise. The power asymmetry of the Western EU nations was supported by EU institutions, highlighted by the relentless search for compromise at the European Commission. Especially in a regional organization as heavily institutionalized as the EU, this support shifts the odds in favor of an initiative.

This case generalizes a divide between the wealthier, Western EU states and the newer Central and Eastern EU states to manage complexity. Nonetheless, significant differences existed within each group, especially due to the significance of reputational factors within the EU. For example, one foreign minister claims that Italy’s economic, demographic, and geographic advantages meant little as the country experienced political instability.131 This demonstrates the potential for error in generalizing a divide between EU nations on the basis of wealth.

CONCLUSION
The EU case shows that power asymmetry is enabling when it drives wealthier states to carry the cost of cooperation for a group. The Western EU member states were willing to shoulder the cost of the 2020 Climate and Energy Package to allow the EU’s climate policy to progress without harming other members. The wealthier nations also benefit from regional climate policy, as it allows their individual climate change efforts to become part of a more meaningful and influential regional climate regime.

The EU case also supports the conclusion that a power asymmetry is beneficial only within a certain range. A moderate as opposed to an extreme power asymmetry ensures that powerful states accommodate the interests of the less powerful members. In the EU case, the wealthier, Western EU states had to accommodate the interests of the Central and Eastern members, exemplified by the concessions Poland was able to negotiate in the name of promoting domestic economic growth. This accommodation promotes satisfaction and in turn compliance with the negotiated agreement. Furthermore, identifying fair criteria to determine emissions reductions, in this case GDP per capita, further benefits future EU climate efforts; it creates the expectation that those states with low GDP per capita will reduce their emissions once their GDP per capita rises. As states become wealthy, like Spain did between 1990 and 2007, they are expected to enact rigorous climate policies. Eventually, as the GDP per capita of Central and Eastern EU states rise, they will also be expected to reduce emissions. This pattern benefits the overall development of EU climate policy.

The moderate power asymmetry in the EU likely led to a cooperative outcome in part because of the EU’s institutional framework. As collective action theory posits, when actors envision a long time horizon for cooperation, they have a greater incentive to work together as they expect continued interaction with one another.132 This creates the need to take everyone’s interests into account. This need is shown in the cooperative mechanisms and procedures of the EU. The Commissions’ structure, which allows member states to appoint Commissioners, ensures that member states can embed their interests within the institution. Other mechanisms, like the rotating presidency in the European Council, give every state access to a powerful platform.133 Similarly, in the ministerial body of the EU, the Council of the European Union, ministers prefer to avoid voting to pass decisions with a majority, and instead work until consensus is achieved.134 This demonstrates that the EU is designed to be a more cooperative than coercive organization.

Finally, the support of powerful EU institutions for climate change policy suggests that a more subtle form of power was at work. As Lukes’ three-dimensional view of power explains, institutions are embedded with power that can unconsciously affect the interests of actors and the outcome of a conflict of interest.135 This was likely the case for the 2020 Package negotiations, as the actors supporting the package had a longer history of EU membership, giving them more time to embed their interests in the EU institutions. Moreover, the clear use of norms (the principle of fairness and the importance of EU climate leadership) were also propagated through the EU institutions. These factors suggest that the Commission had a Western European bias, meaning the institutional environment of the EU favored a climate agreement.

MERCOSUR CASE STUDY: THE 2001 ENVIRONMENTAL FRAMEWORK AGREEMENT
The following case explores the nature of power in
negotiations for the 2001 Environmental Framework Agreement. The negotiations appear to exhibit a conflict of interest between the advanced environmental policies of Brazil and the economic interests of Argentina. The fact that the environmental interest faltered in the negotiations despite Brazil’s status as Latin America’s hegemon raises questions over the nature of power asymmetry in negotiations. One possibility is that Brazil was not invested enough in the outcome of the negotiations to protect its environmental interests, especially given the ferocity of Argentine resistance. This is likely given that Brazil faced an economic crisis in 1997, which distracted the Brazilian government. Another possibility is that high rates of poverty and inequality in Brazil prevent it from exerting economic power, because side payments or carrying the cost of cooperation were not acceptable. These propositions cast doubt on the extent to which Brazil is Latin America’s hegemon.

CONTEXT

MERCOSUR’S DEFINING FEATURES

One of the unique features of regional cooperation in Latin America is the frequent rise of new regional organizations with overlapping purposes. Mercosur is hence the youngest organization in this study, as it was established in 1991 when Brazil, Argentina, Uruguay, and Paraguay signed the Treaty of Asunción. The bloc was noteworthy for bringing together former rivals, Brazil and Argentina, thus uniting Latin America’s biggest and third biggest economy in one trading bloc.

Mercosur was considered one of the region’s most promising integration initiatives. With its economic focus, Mercosur increased intra-regional trade from ten percent in 1990 to 25 percent in 1995. But following the Brazilian currency devaluation in 1999 and the Argentine economic crisis in 2001, both Brazil and Argentina unilaterally reverted to protectionist measures. This showed how Mercosur sought to unite the region, but domestic economic problems brought the return of a competitive mindset.

Mercosur also existed to ensure the bloc’s future relevance. As one Brazilian official described it, in the 1990s they believed that the world was being divided into regions, and that to compete, one had to join this movement. Mercosur thus served as a tool to unite the four countries, creating a stronger platform to engage nations outside the region. The bloc also served Brazil’s individual aspirations. Brazil needed its neighbors to support its leadership as a platform for global power projection. As such, Mercosur was created with global power dynamics in mind.

Another defining feature of Mercosur in contrast to the EU and ASEAN is its reliance on presidents for decision-making. The relevance of the regional organization hence depends on the rise and fall of the presidential administrations invested in them. For example, while the leftist Argentine president Cristina Fernández de Kirchner was committed to the regional organization the Union of South American Nations (UNASUR), her neoliberal successor Mauricio Macri was committed to Mercosur. Macri subsequently dismantled all parts of the foreign ministry responsible for UNASUR and cut Argentine monetary contributions to it. Presidents also tried to transform the organizations’ focus based on their interests. Mercosur thus shifted from a pragmatic trade integration project, to one of progressive political activism and leftist ideologies, and back to one of trade liberalization. The changing presidents weaken organizations by shortening their time horizon and making them susceptible to dramatic changes.

THE PATH TO THE 2001 ENVIRONMENTAL FRAMEWORK AGREEMENT

The responsibility for drafting the 2001 Environmental Framework Agreement was given to the technical sub-committee SGT N6, consisting of representatives from the environment ministries. The national coordinator for the Argentine delegation to SGT N6, Elizabeth Mirta Laciár, describes how SGT N6 began drafting what they believed would become an additional protocol for the Treaty of Asuncion, meaning it would be a legal instrument to orient the harmonization of environmental laws. The SGT N6 consulted with NGOs, private sector entities, and other government bodies for this draft. It was created by the Brazilian delegation, which was the President Pro-tempore of SGT N6. They created an extensive and broad document, reflecting Brazil’s more developed domestic environmental policies.

Yet the SGT N6’s drafts were rejected by the Common Market Group (GMC), which consists of national representatives from the Ministry of the Economy, Foreign Affairs, and Central Banks. The first draft in 1997 failed due to Argentinian opposition to regulating biotechnology and to the influence of the “precautionary principle.” The next draft was submitted to the GMC in March 1999, but the GMC submitted.
did not review it for another one and a half years. The long timeframe demonstrates their lack of interest in the process. Again, the GMC rejected the document due to opposition from the Argentine delegation.\textsuperscript{154}

After failing to produce an acceptable draft both in 1997 and 1999, the SGT N6 realized that the next draft needed approval to ensure their survival as a subcommittee, making the Argentine delegation to the SGT N6 more cautious.\textsuperscript{155} Nonetheless, when the Argentine Ministry of Foreign Affairs and Culture and the Economic Ministry received the draft, they were not satisfied; they stripped it of specific directives and mandates until it merely reiterated past agreements.\textsuperscript{156} By the end, the agreement was a diminished version of the 1992 Rio Agreement—even cutting out the precautionary principle—without the legal status as a protocol.\textsuperscript{157} The process of negotiations thus began with expansive and far-reaching proposals from the environmental ministers in the SGT N6, but was ultimately reduced to the most barebones consensus through the influence of the foreign and economic ministries.

**POWER ANALYSIS**

The following section explores to what extent Mercosur members had a clear conflict of interest, access to power resources, and the capacity to mobilize them for the 2001 Environmental Framework Agreement.

**CONFLICT OF INTEREST**

The negotiations for the 2001 Environmental Framework Agreement demonstrate a conflict of interest between the more developed environmental regulations of Brazil and the economic concerns of Argentina, though the degree to which Brazil was invested in the outcome is unclear.

Brazil favored a comprehensive agreement due to the advanced nature of its domestic environmental measures. Brazil established its environment ministry following the 1972 Stockholm Conference on the Human Environment, almost 20 years before Argentina did similarly so.\textsuperscript{158} As a result, Brazil has more developed environmental institutions and legislation. The Brazilian delegation’s original protocol draft in 1996 was thus detailed and far-reaching.\textsuperscript{159} This stands in sharp contrast to Brazil’s position on other issues, where Brazilian negotiators were usually the strongest advocate for minimal institutionalization.\textsuperscript{160} Brazil’s selective support of higher environmental standards in the face of general opposition to institutionalization underscores the unique value Brazil placed on this topic. Brazil’s support for environmental measures stems in part from the Brazilian businesses, who favored the “harmonization of norms upward,” to make other Mercosur members match their environmental standards and to avoid the risk of becoming less competitive.\textsuperscript{161} Brazil’s domestic legal structure, initial proposal, and business leaders suggest that Brazil was a voice for regional environmental cooperation.

In contrast to those in Brazil, environmental issues in Argentina have less institutional support, partly explaining its clear resistance to the framework agreement. Argentina first created an environmental secretariat in 1991, but it did not hold the same status as a ministry and was later incorporated into the congested Ministry of Social Development.\textsuperscript{162} Without institutional status in the domestic political system, environmental issues likely lacked political support in an intergovernmental organization like Mercosur. Environmental issues also remained second to the demands of free trade. Even the most active environmental secretary María Julia Alsogaray argued that environmental protection should not interfere with the cause of neoliberalism.\textsuperscript{163} As Argentina is one of the world’s leading exporters in genetically modified crops, they were central to vetoing guidelines on biotechnology in the 2001 Environmental Framework Agreement.\textsuperscript{164} Argentina also opposed the precautionary principle, which asserts that states should apply a “precautionary approach,” in situations with serious threats to the environment, even in the absence of scientific certainty.\textsuperscript{165} Opposing a cautious approach shows the extent to which Argentina resisted deeper environmental efforts.

**ECONOMIC POWER**

Economic power played a role in the negotiations insofar as it determined the weight the four Mercosur members gave each other’s positions. Especially given that Mercosur is an economic trading bloc, economic power sets the frame within which the members view one another. Hence, Argentina’s Congresswoman Cornelia Schmidt Liermann, chairman of the House Foreign Affairs Committee, commented in 2018 that Brazil is the clear leader of Mercosur. Even at times where Brazil cannot lead due to internal instability, Argentina will lead according to Brazil’s expected interests.\textsuperscript{166} The following graph shows the GDP of the four Mercosur nations from 1990 to 2005. The graph shows the dramatic difference in economic weight, highlighting why Brazil, and secondarily Argentina,
have much more influence than Uruguay and Paraguay. It also shows the dip in GDP following Brazil’s currency devaluation crisis in 1999 and Argentina’s economic crisis in 2001.

Given Brazil’s large GDP, one would have expected its interest to triumph in negotiations. The fact that it did not suggests that Brazil could not use its economic power. In the context of regional cooperation, economic power can take the form of monetary transfers to neighboring countries or voluntarily bearing the cost of an initiative. Yet Brazil is not positioned to use its wealth in either form, because its GDP per capita is lower than that of Argentina and Uruguay (shown in the graph below).

Andrés Malamud explains that Brazil’s higher rates of poverty and inequality prevent Brazilian authorities from transferring payments or taking on economic costs to regional partners, as this would be portrayed as a subsidy to its wealthier neighbors. If this hypothesis is correct, then Brazil’s wealth does not translate into economic power despite the size. Another explanation is that the depth of the devaluation crisis caused Brazil to be too preoccupied with economic issues to worry about the 2001 Environmental Framework Agreement, meaning that Brazil chose not to exert power to prioritize other issues. Both propositions may be correct in explaining the limited effect of economic size on negotiations.

Military power had little effect on negotiations. Like the EU, the creation of Mercosur made military power less relevant by establishing a peace zone in 1999. Nonetheless, like economic power, military power can give a nation more legitimacy. The graph below compares the CINC Score of the Mercosur nations. As the CINC score considers demographic, industrial, and military indicators to estimate a nation’s national capabilities, it captures the dramatic differences among the members.

Brazil’s clear advantage in military capabilities does not necessarily make it a military power. A Brazilian official argued in an interview that Brazil’s military is unusual for fulfilling more civilian functions, like infrastructure and engineering projects. Brazil also does not have a militaristic history, as its last major war was from 1865 to 1879 when Brazil aligned with Argentina and Uruguay to defeat little Paraguay. In this sense, Brazil does not use its military to project power.

OTHER FORMS OF POWER — NGO ADVOCACY
NGO lobbying played a part in introducing an environmental interest into Mercosur, as NGOs argued for the need to unite environmental and economic interests on human rights grounds. The NGOs thus exerted normative power to the extent that they could shape the conversation on regional integration. This follows Lukes’ third view of power, where actors can affect
the wants of another actor through the nature of the conversation they drive. The NGO interest was led by the Foundation Ecos, which along with fifteen other NGOs from Mercosur nations launched a program on trade and the environment to promote a legal instrument for standardizing environmental regulations. The extensive consultations of SGT N6 show that this interest by NGOs was welcomed and willingly included by the Mercosur nations’ environmental representatives. The former director of the Foundation Ecos, María Leichner, also published a piece largely criticizing one of the SGT N6’s draft protocol. The platform given to her criticism shows how NGOs were an important player in environmental discussions in Mercosur.

THE 2001 ENVIRONMENTAL FRAMEWORK AGREEMENT’S EFFECT

The 2001 Environmental Framework Agreement contains little substance that furthers regional cooperation. Using Keohane’s distinction between harmony and cooperation, it becomes clear that the 2001 Framework Agreement is little more than harmony. The attempt by the SGT N6 to work through conflicts of interest among member states to create meaningful environmental regulations falls flat when the Ministries of the Economy and Foreign Affairs take over negotiations. By simply rehashing past agreements, the 2001 Framework agreement merely exalts the existing harmony rather than resolving deeper conflicts of interest.

THE RESULTING POWER ASYMMETRY

The negotiations leading up the 2001 Environmental Framework Agreement show how the power asymmetry depends upon the level of interest of the involved actors, as this determines the lengths they are willing to go to defend their interest. While Mercosur had a conflict of interest between the well-developed environmental proposals from Brazil and the economic interest from Argentina, the Brazilian government seemed less interested in the outcome of the negotiations than Argentina. As Lukes’s two-dimensional view of power suggests, actors may intentionally decide not to act in anticipation of the other’s response; Brazil may have perceived the intensity with which Argentina opposed environmental regulations and calculated that the resources spent to exert influence were not worth the cost. Moreover, due to the 1999 currency devaluation crisis, Brazil was most likely too preoccupied with domestic challenges to focus on directing resources toward the negotiations. The Brazilian delegation to SGT N6 may have drafted a comprehensive proposal in anticipation of the Brazilian government’s support, but due to the crisis it never received this top level backing. While the exact reason for the absence of top level government support is unclear, what is clear is that without it, Argentina had a stronger interest in blocking an agreement which moved the power asymmetry in its favor.

Another explanation for a power asymmetry against Brazil’s environmental interest is that low GDP per capita limited its influence. As Malamud argued, Brazil’s poverty and economic inequality hinders it from exerting economic power in the form of side payments or in carrying the cost of agreement. This obstacle is especially relevant in the environmental arena, given that environmental interests are often secondary to larger economic or social problems. The course of the 2001 Framework Agreement therefore speaks to the importance of a surplus in economic resources for issues that have a lower political priority.

CONCLUSION

The Mercosur case shows that the power asymmetry does not support cooperation when it favors an actor opposing a cooperative outcome. For the 2001 Agreement to achieve a new impetus for regional environmental standards, it needed Brazil’s backing. Instead, the agreement only received Argentina’s staunch opposition. The power asymmetry in favor of Argentina further shows how regional cooperation requires a leader willing to take the interests of a variety of actors into account. While the SGT N6 did expansive consultations with various actors in developing its proposals, the economic and foreign affairs ministries ignored these interests in writing a new, bare-bones version of the agreement. This case thus shows how the power asymmetry favors a particular type of actor willing to not only protect their own self-interest but act in the interest of the region.

The lack of cooperation may also be the product of the great imbalance in power among Mercosur’s four members, derived primarily from differences in size and material resources. A more balanced spread of power would force stronger members to accommodate some of the region’s other interests. Instead, Brazil and Argentina hold such power that they can easily overlook Paraguay and Uruguay. This encourages the countries to think only of their own interests. In the case of the 2001 Environmental Framework Agree-
ment, the power imbalance may have triggered Argentina to act in its interest with little regard for what is beneficial for the other members. Extreme imbalances in power may have prevented a more collaborative spirit from taking hold in the region.

Finally, the power dynamics did not support cooperation, as the general lack of presidential interest prevented the final agreement from breaking new ground. Mercosur’s dependence on presidential diplomacy means that covering new ground would have required a genuine exchange of interests and attentive negotiation among the presidents. It is possible that this high-level attention was expected, given that the SGT N6 initially made groundbreaking drafts for the protocol, but due to other issues never came. The SGT N6 was soon undercut by the more powerful ministries of the economy and foreign affairs. This is not necessarily surprising, however, as the power of the ministries of the economic and foreign affairs in the Mercosur hierarchy reflects the priorities of its member states. For Mercosur, the 2001 Environmental Framework Agreement was merely symbolic and perhaps intentionally so.

CONCLUSION
This study began with the following question: how does power asymmetry affect regional cooperation? I argued that power asymmetry’s effect depends upon three factors: the existence and depth of a conflict of interest, the power resources available to relevant actors, and the ability of those actors to direct their resources at their targets.

While all three factors play a role in each case, a different one was decisive in each. The depth of the conflict of interest was decisive in the Mercosur case. Brazil’s interest in a regional agreement was too low to incentivize the use of power resources to pressure Argentina. Meanwhile Argentina’s determination to prevent extensive environmental measures motivated a strong stance in opposition, causing a shallow agreement to emerge.

The power resources available to invested actors were decisive in the EU case. The Western EU members were able to drive the development of the 2020 Climate and Energy Package using their economic resources. For one, they were able to carry the major economic costs of the agreement and to embed it with side payments to resistant members. Moreover, they gave the European Commission a bias in favor of climate change policy, creating an institutional supporter that relentlessly developed the agreement until it became acceptable for other member states.

Finally, the ability to direct power resources toward a target was decisive in the ASEAN case. On one hand, Singapore and Malaysia had a high capacity to mobilize and direct resources. Especially Singapore stood out for its ability to not only use its own economic wealth as influence, but to mobilize other actors like the United Nations or the regional intellectual community as a source of influence. On the other hand, the agreement’s effectiveness was limited by Indonesia’s corrupt ministries, where a lack of accountability made the multitude of actors difficult to influence.

The three cases highlight that a moderate power asymmetry benefits cooperation by providing someone to smooth out differences among actors, drive policy innovation, and carry costs associated with an agreement. When there are clashing interests, power asymmetries may give negotiations greater direction to make them more productive.

ANALYTICAL FINDINGS
THE NEED TO PROBLEMATIZE POWER
The primary challenge in this study was not so much identifying whether a power asymmetry supported regional cooperation, but rather identifying whether a power asymmetry existed to begin with. This challenge speaks to power’s multidimensionality, as its many forms produce power asymmetries that favor unexpected actors. Power’s multidimensionality underscores the importance of problematizing power in every situation based on the contextual circumstances and the issues under consideration. Much of the scholarship on regional cooperation nonchalantly assumes that Brazil or Indonesia are regional powers. Yet, the cases in this study illustrate how their respective internal instability prevented them from exerting power. Likewise, scholarship often treats Germany and France as the EU’s defining powers. While these countries were in favor of the 2020 Climate and Energy Package, the extent to which Central and Eastern European nations were able to negotiate concessions underscores the influence of unexpected actors. Power’s dynamism means that the obvious actor will not necessarily triumph, but that unforeseen contextual factors can alter the playing field to produce a surprising outcome.

The question of internal stability raised another fascinating element of power often overlooked
in international relations theory: it is not enough for an actor to have the capacity to exert power, but the target of that exercise of power must have a degree of internal cohesion in order to be influenced. The case of Indonesia and the 2002 Haze Agreement show how corruption in the Ministry of Forest as well as a decentralized government more broadly prevented the Indonesian government from implementing the agreement.\textsuperscript{174} If a government is too disjointed or corrupt to implement a decision, influencing that government becomes harder. Then, the government does not act as a unitary actor, but rather as a multiplicity of actors ranging from powerful individuals to entire ministries relevant to a decision. Influencing these actors is difficult, if, as in Indonesia, they personally benefit from the illegal activity an agreement is trying to prevent. Similarly, albeit to a lesser degree, Argentina’s tug-of-war between the broad ambitions of the delegation sent to prepare the 2001 Environmental Framework Agreement and the resistance of the Argentine Ministry of Foreign Affairs and Culture and the Economic Ministry show the divisions that make it difficult to treat a country as a unitary actor. Yet, influencing a cacophony of domestic actors is challenging, especially because the precise nature of internal political battles is hard for an outsider to gauge.

THE VALUE OF AN ACTOR DRIVING NEGOTIATIONS
All three cases highlight the importance of an actor’s role in driving negotiations. This is particularly important in developing a proposal acceptable to all parties in the negotiations. As collective action theory suggests, collective action becomes more difficult with an increasing number of actors, as the presence of more players encourages free-riding and raises the transaction costs of reaching an agreement.\textsuperscript{175} Due to the EU’s large size, the Commission was essential in absorbing these transaction costs through a relentless pursuit of agreement and in countering the free-riding that would have undermined one. For less institutionalized and more intergovernmental organizations like ASEAN and Mercosur, member states have to take on this driving function. Here, the state’s capacity to effectively mobilize resources becomes invaluable. This capacity enhances Singapore’s influence in ASEAN, allowing Singapore to not only assert its interests in the negotiations of the 2002 Haze Agreement, but also in a variety of other ASEAN initiatives. In contrast, Mercosur’s cooperation efforts suffer from the absence of a successful policy driver. In fact, Latin American nations generally struggle due to internal instability and corruption, meaning few countries can consistently take on this driving role. This may be one explanation for the short lifespan of regional organizations in Latin America.

In addition, actors are able to drive negotiations forward by absorbing their costs. As hegemonic stability theory posits, a hegemonic actor benefits cooperation in instances where the powerful actor is willing to carry disproportionate costs of an institution. In both the EU and ASEAN cases, the countries in favor of a regional agreement were willing to carry greater economic costs in driving negotiations. In the EU this meant that wealthy, Western nations were primarily responsible for reducing emissions, while in ASEAN this was shown in the monetary investments of Singapore and Malaysia in driving the content development of the 2002 Haze Agreement. Yet, this willingness to carry costs is clearly absent in Mercosur. Brazil’s inability to take on this leading role stems in part from its domestic economic weakness. Alternative explanations see Brazil as less invested in the regional platform due to the belief that it is already a great power on its own. Regardless of its underlying reasons, Brazil’s inability to take on a leadership role in Mercosur that takes other actor’s interests into account explains the lack of development in environmental policy.

ONLY A MODERATE POWER ASYMMETRY SUPPORTS CO-OPERATION
While the likelihood that an agreement will emerge rises when a country or institution champions an agreement, the quality of emerging cooperation depends upon the ability of this champion to incorporate other actors’ interests. In the EU, even countries as small as the Czech Republic played a role in developing the 2020 Climate and Energy Package. Moreover, the EU agreement was crafted with the principle of fairness in mind, meaning wealthy nations willingly took on greater costs to carry the agreement. Embedding many members’ interests in the agreement ensures a certain degree of satisfaction and support, increasing the likelihood of the agreement’s successful implementation. However, in the Mercosur case, the interests of Paraguay and Uruguay did not play a role in negotiations. The extreme power imbalance within Mercosur has allowed Argentina and Brazil to continuously ignore the organization’s smaller members. This has produced such dissatisfaction that taxi driv-
ers passing the Mercosur headquarters in Uruguay’s capital need little provocation to go on extended rants regarding the organization’s oppressive nature. Likewise, regional cooperation was hurt by the lack of burden sharing in ASEAN’s 2002 Haze Agreement, where the costs almost entirely fell upon Indonesia. This gave Indonesia little incentive to comply with the agreement. In sum, the cases support the critique of hegemonic stability theory in that an extreme power asymmetry undermines long term cooperation by allowing the champions of an agreement to ignore and overrule the interests of the other actors.

THE FUTURE OF REGIONAL COOPERATION
As international relations scholars speak of the “regionalization” of global politics, the many factors necessary for states to work together show that regionalization will not necessarily take shape in a cooperative fashion. These case studies highlighted the need for particular countries to drive regional initiatives forward. Given power’s multidimensionality, even small states have roles to play in supporting cooperative outcomes. This was shown in the EU, where small states were at the forefront of leading climate change policy through their own example and expertise, as well as in ASEAN, where Singapore and Malaysia united the other ASEAN nations to pursue an agreement.

The cases also highlighted the importance of economic prosperity for the future of regional cooperation. In both ASEAN and the EU, wealthy states were able to take on the costs of regional cooperation. Given that Brazil, with its low GDP per capita, plays less of a leadership role in Latin America, one may infer that the wellbeing of a country’s citizens supports a quest for leadership. If this is true, then the success of regional cooperation will also depend on the future economic wellbeing of societies. This uncertainty potentially motivates policymakers to build the foundations for cooperation now, so that states will be able to support each other’s mutual prosperity in the future.

Endnotes
5 Ibid, 53-54.
9 Ibid, 19.
10 Ibid, 27.
11 Ibid, 17.
12 Keohane and EBSCO Publishing (Firme), After Hegemony, 57.
17 Lukes, *Power*.
18 Ibid.
22 Ibid, 43.
31 Ibid, 193.
32 Ibid, 196.
33 Cited in Ostrom, 197.
34 Ostrom, 202-03
38 Lukes, *Power*.
40 Lukes, *Power*, 27.
47 Lukes, *Power*.
51 Koh and Robinson, 4.
53 Author’s Interview with Ibrahim Almuttaqi, Head of the ASEAN Studies Program at the Habibie Center, August 1, 2018 in Jakarta, Indonesia.
54 Author’s Interview with an ASEAN Official, August 2018.
56 Ibid, 2.
57 Ibid, 79.
59 Nguitragool, 79; Daniel Heilmann, “After Indonesia’s Ratification: The ASEAN Agreement on Transboundary Haze Pollution and Its Effectiveness As a Regional Environmental Governance Tool,” *Journal of Current Southeast Asian Affairs*, 34, no. 3 (2015), 102.
60 Heilmann, 102.
62 Ibid, 81.
63 Ibid, 83.
64 Ibid, 70.
65 Heilmann, “After Indonesia’s Ratification: The ASEAN Agreement on Transboundary Haze Pollution and Its Effectiveness As a Regional Environmental Governance Tool,” 97.
67 Heilmann, “After Indonesia’s Ratification: The ASEAN Agreement on Transboundary Haze Pollution and Its Effectiveness As a Regional Environmental Governance Tool,” 101.
74 Ibid, 65.
75 Ibid.
76 Ibid, 82.
78 Nguitragool, 127.
82 Mahbubani and Sng, *The ASEAN Miracle*.
83 Author’s interview with a Singaporean Diplomat, July 30, 2018.
86 Ibid, 89.
87 Jerger, “Indonesia’s Role in Realizing the
Goals of ASEAN’s Agreement on Transboundary Haze Pollution,” 36.


89 Allison and Halperin, “Bureaucratic Politics: A Paradigm and Some Policy Implications,” 70.

90 European Commission, “From 6 to 28 Members.”


92 Delreux and Happaerts, Environmental Policy and Politics in the European Union, 59.


99 Ibid, 695.

100 Skjærseth, “Unpacking the EU Climate and Energy Package: Causes, Content and Consequences,” 17.

101 Delreux and Happaerts, Environmental Policy and Politics in the European Union, 216.

102 Skjærseth, “Unpacking the EU Climate and Energy Package: Causes, Content and Consequences.”


107 Delreux and Happaerts, Environmental Policy and Politics in the European Union, 217.

108 Delbeke and Klaassen, “Framing Member States’ Policies,” 93.


111 Braun, “EU Climate Norms in East-Central Europe,” 450.

112 Ibid, 450-51.


115 Delreux and Happaerts, 31; Schreurs and Tiberghien, “Multi-Level Reinforcement: Explaining European Union Leadership in Climate Change Mitigation,” 32.

116 Braun, “EU Climate Norms in East-Central Europe,” 454.


120 Skjærseth, “Unpacking the EU Climate and
Energy Package: Causes, Content and Consequences,” 28.
121 Delbeke and Klaassen, “Framing Member States’ Policies,” 93.
122 Kupchan, How Enemies Become Friends, 2.
124 Ibid. 
125 Ibid.
128 Delbeke and Klaassen, “Framing Member States’ Policies,” 93–94.
130 Delreux and Happaerts, Environmental Policy and Politics in the European Union, 221.
132 Ostrom, “Collective Action Theory.”
135 Lukes, Power.
141 Mario E. Carranza, “La Institucionalidad 108
145 Author’s interview with an Argentina Diplomat, June 6, 2018 in Buenos Aires, Argentina; Author’s interview with Professor Alejandro Frenkel from the Universidad Nacional de San Martin, June 5, 2018 in Buenos Aires, Argentina.
146 Malamud, “Interdependence, Leadership and Institutionalization: The Triple Deficit and Fading Prospects of Mercosur.”
149 Ibid, 94.
150 Hochstetler, “Race to the Middle: Environmental Politics in the Mercosur Free Trade Area,” 352. 
151 Ibid.
153 Hochstetler, “Race to the Middle: Environmental Politics in the Mercosur Free Trade Area,” 352; Laciar, “El Instrumento Jurídico Ambiental Del Mercosur,” 96.
154 Laciar, 97.
156 Laciar, 100; Hochstetler, “Race to the Middle: Environmental Politics in the Mercosur Free Trade Area,” 353.
157 Hochstetler, 353.
158 Kathryn Hochstetler, “Fading Green? Envi-

159  Ibid, 11; 19.

160  Ibid, 19.

161  Hochstetler, “Race to the Middle: Environmental Politics in the Mercosur Free Trade Area,” 352.


163  Ibid, 10.


166  Author’s interview with Cornelia Schmidt Liermann, June 5, 2018 in Buenos Aires, Argentina.


170  Author’s interview with a Brazilian official knowledgeable in security matters, June 19, 2019 in Rio de Janeiro, Brazil.


176  Author’s experience in Montevideo, Uruguay in June 2018.

Bibliography


Carranza, Mario E. “La Institucionalidad ‘Ligera’ Del Mercosur y Sus Perspectivas de Sobrevivencia


Studies 41, no. 01 (February 2009): 27. https://doi.org/10.1017/S0022216X08005105.


UNGA. “Report of the United Nations Conference on


Primary Sources—Author’s Interviews

Author’s Interview with Ibrahim Almuttaqi, Head of the ASEAN Studies Program at the Habibie Center, August 1, 2018 in Jakarta, Indonesia

Author’s Interview with an ASEAN Official, August 2018

Author’s interview with a Singaporean Diplomat, July 30, 2018

Author’s interview with a Brazilian Civil Servant, May 2018

Author’s interview with an Argentina Diplomat, June 6, 2018 in Buenos Aires, Argentina

Author’s interview with Professor Alejandro Frenkel from the Universidad Nacional de San Martin, June 5, 2018 in Buenos Aires, Argentina

Author’s interview with Cornelia Schmidt Liermann, June 5, 2018 in Buenos Aires, Argentina

Author’s interview with a Brazilian official knowledgeable in security matters, June 19, 2019 in Rio de Janeiro, Brazil

Primary Sources—Data Sets


The Future of Economic Geopolitics: Network Effects in Intercultural Trade

Joshua Benjamin Curtis, Duke University Class of 2019

Abstract

Using a regression discontinuity design on a gravity model of trade among 36 Middle Eastern and East Asian countries between 1980 and 2014, this study demonstrates network effects in trade. A small increase in trade between subsets of two cultural blocs diminishes the effect of cultural similarity on trade between all members of the two cultural blocs. The result holds regardless of whether cultural similarity was originally a boon or drag on trade. Furthermore, international businesses adjust to new intercultural acumen very rapidly. The effect demonstrated herein points toward an answer to economic dilemmas posed by Huntington’s “clash of civilizations.”

International trade is an important part of international economic development. It permits individuals in countries with more trade the ability to choose from a greater variety of goods. It furthermore allows firms to sell their goods to a larger market at potentially higher prices, thereby sustaining business when domestic demand cannot. These firms become more productive through broader business contact. Moreover, trade leads to competition, as the most productive firms compete internationally, absorbing customers and resources from the least productive firms. This tendency does not only improve productivity. It also weakens monopolists’ power to set high prices on goods like cereals or telecommunications, and power to force down wages as foreign firms set up operations locally—both developments that especially help the poor in developing countries.

Nevertheless, it has historically been fashionable to claim that trade globalization benefits Western elites at the expense of the developing world. On the contrary, Western countries stand to lose the most economic influence from globalized trade going forward. Even though the combined GDP of Western countries in 2014 was over $40 trillion, larger than the combined GDPs of all other countries by only about $3 billion, trade among Western countries clocked in at only $5.6 trillion, well below the $6.1 trillion traded among non-Western countries. Currently, many countries look to the robust, advanced economies of the West for business opportunities, yielding an additional $6.5 trillion. However, this situation will change rapidly, as non-Western economies strengthen and deepen. By 2050, analysts at PwC project that the economies of non-Western countries like China, India, Indonesia, Brazil, Russia, Mexico, Turkey, Nigeria, and the Philippines will easily dwarf Western economies as they stand today. Meanwhile, the growth in international trade in the late 20th century was already shown to come mostly among countries that shared key cultural characteristics. Although interregional trade grew and deepened over that timeframe, it was outsourced by trade within regions such as East Asia or...
Latin America. Cultural barriers therefore threaten not only to temper global trade between regions and the various benefits derivative thereof, but also particularly threaten to sideline Western business as trade among non-Western growth centers accelerates in comparison.

There is also a concern—particularly in light of China’s recent trend of exporting its police-state technology and supporting friendly dictators and corrupt regimes—that not all trade relationships should be equally encouraged. The decline of economic influence of Western countries thus has far-reaching impacts.

While this perspective on globalization—based on regional competition—may be a new approach in economics, it has deep roots in other disciplines. In the wake of the Cold War some political scientists presumed that globalization would make the world unambiguously freer, closer, richer, and more stable. In contrast, Samuel Huntington famously predicted instead that the world would return to older divisions based on broadly-shared cultural customs and histories, a theory now dubbed the “clash of civilizations” (see Figure 1).

Huntington’s theory has drawn vocal criticism. One camp argues that he simply repackages geopolitical realism—the idea that international relations must derive from zero-sum great power competition—as a historical fait accompli with political science jargon to increase its authority. This camp demonstrates that the theory is in fact not realistic at all, as it fails to accurately predict violent conflict. Although the strongest of these criticisms were levelled soon after the end of the Cold War and thus had limited data to support their criticisms, they do hold some merit. Formal international relations and conflict have not been primarily reformed along “civilizational” lines in the past 30 years. Many of the most vicious conflicts have erupted between groups within a “civilization” (e.g. Rwanda).

Another criticism asserts that Huntington’s framework of civilizational blocs, whether in contest or not, is fundamentally flawed. Huntington, Said posits, downplays the dynamic nature of civilizations, taking their present “boundaries” and their very natures as historically rigid, exogenous determinants of future conflict rather than pliable products of highly intertwined histories. Some go so far as to say that Huntington perhaps intended to stoke, rather than merely predict, a clash of civilizations. In light of the many Western partisans preaching civilizational warfare, Huntington’s work might be read as a self-fulfilling prophecy at best, or as incitement at worst.

Both sets of criticisms, however, fail to convincingly debunk Huntington’s central contention: culture, from local to regional levels, matters. Clearly, inter-civilizational “fault lines” do coincide with many conflicts (e.g. Yugoslavia, Kashmir, Ukraine, Xinjiang/East Turkestan, the immigration debate in the U.S.). Seminal studies by Melitz, Toubal, Helble, and others, discussed at great length in the literature review, show that broad markers of cultural similarity, such as religion and language, powerfully influence economic relations. Gokmen’s recent, comprehensive econometric study indicates ironically that whatever the fate of Huntington’s predictions for the security realm, his theory holds for the realm of trade. As discussed earlier in this section, the stakes for the West of an economic clash of civilizations are high. The consolidation of regional cultural economic blocs could, when unchecked, seriously dampen the future expected benefits of globalized trade—particularly though not exclusively in the West. Perhaps most importantly, the clash clearly seems to exist in political discourse around the world, whether in the form of white supremacy, dreams of renewed Chinese hegemony, pan-Arab or Islamist visions of a homegrown renaissance, or pan-Africanism. Cultural “x-factors” seem to subtly shape the economic and social dynamics influencing international politics.

In the face of evidence that culture plays an increasingly important role in not only international politics but also economics, it would be wise to under-
stand the nature of global cultural influences on trade and investment. Huntington’s theory is the foundational theoretical model of cultural geopolitics and as such, despite its myriad shortcomings, provides a rough framework to defenders of the international liberal order as they explore challenges facing the global economy.

Using a loose and innovative approach, a workaround to the increasing economic alienation between cultural blocs becomes apparent. Areas with “split loyalties” or that have ties to multiple cultural blocs—which Huntington predicted would be basket cases in the security realm, torn apart by civilizational friction—can perhaps instead serve as “intercultural bridges” in the economic realm. In such a scenario, small improvements in intercultural economics would have non-linear effects. These would be economic conduits between cultural blocs reminiscent of geographic chokepoints (e.g. the Suez Canal or the Strait of Malacca) familiar to students of geostrategy, except relying on the chokepoints’ good relations with those blocs they connect rather than physical or political control. The key to unlocking a peaceful and prosperous future, and avoiding the gloomy isolation feared by Huntington and his would-be successors, would then seem to be a self-conscious, deliberate effort to build bridges, rather than walls, between civilizations.

Policymakers need quantitative rigor to form well-considered strategies for this new economic landscape. If the role of cultural barriers in shaping trade patterns will resemble the role of geographic barriers, the best model of international trade will account for both forms of barrier. The gravity model, as the most successful and versatile model in international economics, fits this role well. It was first developed to model the economic reality that market size and physical distance, with great consistency, are the best empirical predictors of trade patterns. Now often used to model cultural impacts, the gravity model can once again provide much-needed clarity on international economic realities.

In this study, the gravity model of trade is combined with a regression discontinuity design employing data from the Middle East and East Asia during the modern era to ascertain the existence of network effects in intercultural trade. The findings suggest tentative but strong support for the hypothesis that cultural networks affect international trade, leading to non-linear returns to trade across cultural barriers. The paper will continue as follows: Section II will review the literature on the gravity model’s origin and robustness, on its extensions to topics of geopolitical interest, and on indications of network effects in intercultural trade. Section III will outline the theoretical framework. Section IV will describe the collection and cleaning of data as well as the underlying methodology. Section V will contain the results and discuss their interpretations. Section VI will conclude and describe various implications.

LITERATURE REVIEW

THE GRAVITY MODEL

The gravity equation itself is quite old: it was invented by Sir Isaac Newton to describe the force attracting all mass together. In the 1940s, John Q. Stewart, a social physicist at Princeton University fond of drawing parallels between physical and social concepts, began describing a “demographic gravity” model, in which individuals operate like molecules, with relative importance, distance from each other, and a “demographic energy” between them. He indicated that, like in Newton’s gravity equation, demographic energy should be directly proportional to each individual units’ mass, but inversely proportional to the distance between them. Economists picked up the idea and began thinking about its applications.

In 1960, Savage and Deutsch devised an econometrically sound method for investigating oddities in international trading patterns. They showed that, with the simple use of a matrix of export relationships between sets of countries, a null model of trade could be predicted from each country’s prominence in the global economic system (i.e. from its size). This model was the first attempt to point out reasons for patterns as they are. The authors were among the first to question what seemed to be an odd trend of clusters of trade relationships in the Organization for Economic Cooperation and Development (OECD), where comparative advantage did not seem to be the primary motivator. Tinbergen published a book soon after, suggesting that the international economic order revolved not just around large countries but around neighbors and proximate countries. Toward the end of the decade, Armington discovered that countries exhibited a persistent “home bias” in their consumption of goods on the international market despite the predictions of the most rigorously proven theories of trade.

The next few decades saw an emerging consensus that the “social physics” of gravity applied remarkably well to the patterns of international trade:
relationships clustered around nearby countries but skewed toward those countries that produced more per capita. The linear regression derived from the gravity equation’s multiplicative form related the logarithm of trade volume between two countries to the logarithm of each country’s real GDP per capita and to the distance between them. Economists found it to correlate heavily with observed trading patterns, and thus began regularly using it to test theories. In one example, a modified gravity equation was used on the American paper and paperboard industry to ascertain the strength of the theory of intervening opportunities. This theory states that it is not size of and distance between termini, but the availability of opportunities along the way, that determines the path of an economic pattern. Gravity explained 43 percent of the paper industry’s trade, while intervening opportunities explained only 3 percent.22

Nevertheless, mainstream trade theorists had trouble taking seriously the gravity theory’s explanation of trade in terms of the distance between and sizes of economies. It was statistically accurate, but there was no economic theory guiding its interpretation, and there could certainly be no causal links determined using it. In 1979, Anderson derived a theoretically-grounded gravity model from the GDP-based international expenditure equation, assuming identical homothetic preferences across countries for multiple commodities flowing in all directions and constant elasticity of substitution (CES) indifference curves. However, gravity’s “use [was] at the widest limited to countries where the structure of traded-goods preference [was] very similar and, subsidiarily, where trade tax structures and transport cost structures [were] similar.”23

With some more theoretical rigor, gravity models began entering the mainstream. Daniel Trefler custom-tweaked the Heckscher-Ohlin-Vanek model to address its age-old inability to accurately predict trade based on factor abundance.24 His paper demonstrated that none of these amendments fixed the “case of the missing trade” except for allowing variation by productive technology in wages across countries and for including “home bias,” as first suggested by Armington and increasingly confirmed by the gravity equation.25 In 1995, McCallum’s “border puzzle” shocked the trade world. Based on the gravity theory, the case study of Canada and the U.S. demonstrated that without the porous Canada-U.S. border blocking the trade between provinces and states, Canadian interprovincial trade’s current twenty-fold edge on trade with the U.S. would evaporate entirely.26 Included in the paper was a rather striking visual representation of the logic of the gravity equation (Figure 2).

![Economic map of North America](Figure SEQ Figure * ARABIC 2: Economic map of North America)

The immediate shock of that result forced the economics community to wrestle with the remaining problems of the gravity equation: The model was far better at producing valid predictions than the best trade theories, and it easily produced food for thought for economists and policymakers alike (who in the 1990s were very interested in the economic impact of softening borders and were thus shocked at the impact of such an innocuous one as the U.S.-Canada border). Nevertheless, gravity’s results were untrustworthy on account of its “dubious theoretical heritage” and weak assumptions undergirding its mathematical validity.27

In 2004, Anderson and van Wincoop solved the then notorious McCallum border puzzle, deriving the gravity equation from a general equilibrium in which all trade relationships and price levels must be collected to determine each country’s “multilateral resistance” to imports and exports. They determined that the twenty-fold shift in Canada’s trade could not be definitively attributed to the US without a border, because that would reconfigure the entire world trading system, leading to trade deflection and price changes around the world.28 This formulation, where at the very least importer and exporter fixed effects are a necessity for unbiased estimation of trade patterns
using gravity, has become a gold standard in the now rather theoretically rigorous subfield of trade theory devoted to variations of the gravity equation.

The gravity theory has assembled a long-resume of accomplishments. Among them, one study focusing on trade in 1958 to 1988 showed that a 50 percent cut in the markup of goods at a receiving port compared to the departure port in OECD countries led to an 8 percent increase in world trade. Another study showed that decreasing infrastructure quality to the median from 75th percentile levels increases trade costs by 12 percent. The third study demonstrated that political détente significantly improves trade, while the fourth one found that a one day increase in a good’s travel time decreases the chances of that good’s purchase by as much as would a 2.3 percent tariff.

It is now well-known that there are numerous varieties of trade costs that can represent distance in gravity equations. These include the natural ones imposed by the cost and time of international shipping, artificial ones involving tariffs and non-tariff barriers, and perhaps too the ones tied to concepts of information and trust barriers.

**EXTENSIONS OF THE GRAVITY MODEL**

Researchers have extended the gravity model to numerous demographic and international relations concepts that often relate to or proxy for the informational and insurance costs less readily measured than simple geographic distance or tariff levels. One of the first to study such concepts was Melitz, whose 2002 study utilizing gravity as a template indicated the importance of common language, translation, and even simple literacy to trade volumes. Melitz and Toubal later expanded this template to show that common official language, common spoken language, common native language, and even linguistic similarity of linguistically distinct people positively affect trade by facilitating communication and trust. Meanwhile, Casella and Rauch derived a theory, based on the historical and modern experience of many diasporas ranging from overseas Chinese in modern Southeast Asia to medieval Jews and Armenians, whereby a minority with access to “complete information” on foreign markets and instinctive trust due to ethnic ties can improve international trade ties both for themselves and for the system as a whole. The findings indicate that information on culture is an important determinant of international trade.

An empirical study of trade between Niger and Nigeria—which share the Hausa ethnic group on either side of an arbitrary border drawn by the British and French—represents a natural quasi-experiment where researchers could observe the impact of the border on millet and cowpea prices in the presence of ethnic similarity and difference. The researchers found an unequivocal link between ethnic and price difference and speculate in their case that the reason may relate to trust and customs, since many traders in the region rely on short-term informal loans where the implied interest rate and maturation period differs from one ethnic group to the next.

In a groundbreaking 2007 study, “Is God good for trade?”, Helble found, even after controlling for political regime types and conflicts, that religious similarity usually promotes trade. The author utilized a gravity model and innovative measurements of religious similarity between countries—as well as measures of interfaith contact—to highlight the trade-friendly attitude of Christianity, Islam, and, especially, Judaism. This pattern held even while interfaith trade suffered between the former two and Hinduism. Another groundbreaking study on the far-reaching consequences of cultural systems found that protection for external and non-concentrated investments varies systematically across legal systems’ cultural basis via Roman civil law (through France), Islamic law, or English, German, or Scandinavian common laws. These protections have significant effects for international trade, even where a country’s legal system is exogenously determined due to the vicissitudes of imperial history.

Meanwhile, Dunlevy and Hutchinson described three mechanisms by which immigration promotes trade, at least in the history of the United States: immigrants have new preferences, necessitating new trade flows and offering new business opportunities for those providing them; immigrants know about foreign markets and cultures, and often bring in new expertise; finally, immigrants tend to possess transnational networks. All three mechanisms, and indeed most mechanisms by which language, ethnicity, or religion affect trade, can be described as derivations of the informal information and insurance costs cultural barriers impose on international trade.

Since then, most gravity studies address multiple dimensions of “proximity,” such as political, ethnic, religious, linguistic, or historical similarity, including at least one measure for each category. Zhou sought to clarify the rate of change of trade and
determinants of trade patterns since the Cold War’s end. It is a good example of modern gravity specifications that include a bevy of additional variables. These variables include membership in inter-governmental organizations like the International Monetary Fund; former colonial ties, such as that between two countries formerly colonized by France or between France and a former colony; and many others. One especially innovative study by Gokmen included these demographic variables as controls as it confirmed the existence of a “clash of civilizations” in trade via a regression discontinuity design (RDD) across the end of the Cold War. It confirmed the waxing significance of cultural difference as a barrier to trade, especially among former strategic allies no longer bound by Cold War geopolitics.

As far as the effect of politics and policy on trade is concerned, Rose sought to test whether the foreign service—which in most countries sees trade promotion and facilitation as its core objective—is actually effective in promoting and facilitating commerce. He found a positive but non-linear, diminishing impact of embassies and consulates on trade, with the impact of an additional foreign mission in a foreign country increasing bilateral trade by an average of 6 percent. In contrast, Acemoglu and Yared indicated that as multipolar political competition increased immediately following the end of the Cold War, trade volumes diminished not only due to a country’s own militarization, but even simply due to neighbors’ militarization. These countervailing forces demonstrated the importance of stability, diplomacy, and political ties to international trade.

NETWORK EFFECTS

At the same time that economists have increasingly understood how cultural barriers—or more properly, perhaps, the lack of cultural proximity—impose costs on trade, the relative significance to international business of these implicit information and insurance costs has increased. The advent of containerization significantly reduced transportation costs by integrating land and sea transport systems, speeding transfer of cargo by up to 40-fold, and decreasing damage and theft of goods in comparison to the previous break-bulk method of shipping. Since infrastructure plays a crucial role in bilateral trade via transportation costs, increasing development of global infrastructure, well-represented by, though not limited to, China’s Belt and Road Initiative, has increased trade and diminished the marginal impact of new infrastructure and distance on trade. At the same time, communication lines have been shown to significantly influence trade volumes, with increased and dispersed internet service linked to higher trade volumes.

In this Information Age, where overall trade volumes and basic market access are increasing, the ability to enter markets seems relatively less important than the ability to properly interpret information about foreign markets, parlay this information into smart supply chains and marketing campaigns, and develop a trustworthy foreign business reputation. Many factors would impact these capabilities, but the most geopolitically prominent among them should seem to be culture and related demographic factors.

However, it remains unresolved whether the insights into the demography of trade from the last 20 years point to meaningful policy strategies to improve global trade that do not rely on diminishing cultural distinctness itself. The key lies in how international businesses handle their lack of information about foreign business environments, how they obtain this information, and how they act once they have gathered it.

THEORETICAL FRAMEWORK

This study builds on the gravity equation. The most basic formulation of the gravity model relates trade flows to the size of country pairs’ economies and the distance between them as follows:

\[ X_{ijt} = \alpha_1 Y_{it}^{\alpha_2} Y_{jt}^{\beta} \frac{1}{d_{ij}} \]

where \( X_{ijt} \) is the current average value of exports in USD between countries \( i \) and \( j \) at time \( t \), \( Y \) is current real GDP per capita, and \( d \) is any variable representing distance between countries. To estimate the equation, the natural logarithm is taken, yielding:

\[ \ln \ln X_{ijt} = \alpha_0 + \alpha_1 \ln Y_{it} + \alpha_2 \ln Y_{jt} - \beta \ln \ln d_{ij} + e_{ijt} \]

where \( \alpha_1 \) and \( \alpha_2 \) represent the trade elasticities of real GDP per capita for countries \( i \) and \( j \), respectively, while \( \beta \) represents the elasticity of any distance variable. While the theory undergirding the gravity equation is flexible in permitting multiple distance variables, these variables must be readily interpretable and, ideally, should be continuous, logarithmic measures of distance or proximity in order to fit with the mathematical specification above. These needs must
be weighed against a desire to let data speak rather than transforming it unnecessarily.

A theoretically rigorous gravity equation, however, must control for multilateral resistance. This has been shown numerous ways on multiple occasions since the original proof in Anderson and van Wincoop and is best summarized in Head and Mayer, Section 2.47 This is most commonly done with one of many varieties of fixed-effects schemes; this paper is no exception and will outline the approach in the Methodology subsection.

It is well-known that companies, and especially large corporations, tend to seek out business opportunities outside their country, but must weigh the benefit of internationalizing—whether by trade or by investment abroad—against the costs and risks of international business. Standard trade costs are a relatively predictable matter, but implicit costs stemming from communication and trust issues (which create informal information and insurance costs) can pose a risk to international business that many firms may not be willing to undertake. Cultural barriers create significant information costs to international trade, but once the metaphorical ice has been broken by business partners from two separate cultures, firms from countries in each cultural group become more confident operating in countries within the other cultural group.48 Barkema and Drogendijk’s description of internationalization strategy attests to this theory on a more limited scale: Dutch companies profited from large-scale exploration of Central and Eastern European countries, quickly learning from mistakes to enable massive expansion of foreign market share and operations before cautious competitors can scale up their limited, risk-averse operations.49 However, only companies with prior international experience could manage this, and the more remote—measured holistically, not just geographically—a foreign market, the weaker this effect was. Thus, more culturally distinct markets encourage slow growth, but events that bring large knowledge transfers between international markets and facilitate knowledge transfer among internationally-operating firms can lower the risk of expansion enough that activity explodes. It is worth extending the idea beyond firm specific interactions to the aggregate behavior of firms engaging in international trade.

This study proposes that, while all trade costs are exogenous to individual firms, demographic trade costs are much more sensitive to the status of aggregate trade patterns than either government policies or natural trade costs. Specifically, there are significant third-country effects of intercultural trade for countries that share a culture. As a result, a substantial increase in trade between a subset of countries in culture group A and countries in culture group B should lead to an increase in overall trade between country pairs in culture groups A and B. If true, this would be a departure from the standard assumptions of gravity models that decisions on the size and destination of each consignment of goods are independent.50

**DATA AND METHODOLOGY**

This study uses a multidimensional gravity model, on account of its theoretical and empirical rigor, with a Regression Discontinuity Design (RDD) to demonstrate cultural trade network effects in a limited case study of the bilateral trade flows within and between countries belonging to two cultural blocs. The model estimates trade relations between countries within and between the two regions across a key, singular, exogenous improvement in the recent trade relations of a relatively small subset of countries in one region with countries in the other. The RDD demonstrates that the effect of a limited change in intercultural economic relations reaches beyond the countries directly to trade within and between both cultural blocs; the method also forces the scope to exclude countries that are not in either region as the RDD can only effectively divide across the event for the cultural blocs involved in the discontinuity event. The observations represent panel data on country-pairs across time, with the “dyad-years” covering all unique permutations of two distinct countries for a pool of thirty-seven countries over the thirty-five years from 1980 to 2014.

**DATA COLLECTION**

The study addresses trade within and between the “greater Middle East” and “greater East Asia.” The “greater Middle East” consists of the Arab League, Turkey, Iran, Afghanistan, and Pakistan, while Huntington’s “Sinic,” “Japanese,” and “Buddhist” civilizations—consisting of China, Mongolia, South Korea, Japan, Taiwan, Vietnam, Laos, Cambodia, Thailand, Myanmar, Bhutan, and Singapore—are agglomerated into “greater East Asia.”51 Palestine, Israel, Yemen, and North Korea were excluded due to the unique nature of their geopolitical history in this timespan and/or the unavailability of reliable data. This yields a total of thirty-six countries.

These categorizations are based on versions
of Huntington’s theory amended for a combination of historical reasons. Huntington’s treatment of all Muslim countries as a singular “Islamic civilization” seems unduly influenced by an orientalist understanding of Islamic peoples: Middle Eastern countries, some of which include or recently included large non-Muslim populations, share ancient cultural and historical legacies dating to long before the rise of Islam, while countries such as Indonesia were until quite recently Hindu and Buddhist, with strong ties to a multiplicity of other eastern and southeastern Asian countries.\(^{52}\) The “Middle East” considered in this paper, in contrast, has been closely linked by various empires and political movements, particularly in recent centuries, and many of its countries collaborate economically, socially, and/or politically. This “greater Middle East” is commonly analyzed together by the Central Intelligence Agency, State Department, and U.S. military, as well as both Western and Middle Eastern news outlets.

On the other hand, it is more difficult to delineate sensible boundaries between the distinct but intertwined cultures of China, Korea, Japan, and Southeast Asia, which spent millennia developing common ideas and sharing formative histories largely separate from those of the West. Buddhism, Daoism, Confucianism, and other such “eastern” religions and philosophies are regarded together partly because they borrowed so heavily from each other while having much less contact with others throughout their histories.

The Middle East’s persistent underdevelopment and geopolitical importance, moreover, makes a study of its trade relations particularly attractive. On the other hand, East Asia’s economic prominence has grown significantly since 1980. This allows new international economic connections to develop, permitting better insight into network effects in trade than would a study involving the West, which for decades prior to this study’s scope had strong economic ties with underdeveloped regions such as the Middle East.

Starting after the oil crises of the 1970s, the containerization revolution, the elimination of the Bretton Woods system, the end of Cultural Revolution in China, and the unification of Vietnam, and including in its timeframe the acceleration of global economic liberalization, the late and post-Cold War, and the rise of the Information Age, the 1980-2014 timeframe fairly represents the modern era. At the same time, it does not represent an era of economic homogeneity. With thirty-six countries, thirty-five years, and excluding duplicate dyad-years (e.g. Singapore-Iran-1986 and Iran-Singapore-1986), the remaining dataset contains 22,050 observations.

The methodology derives from a straightforward extension of the gravity model that cautiously includes cutting-edge variables measuring physical, political, and cultural proximity, while avoiding diluting explanatory and interpretive power with either a “kitchen sink” approach or excessive deviation from the mathematical theory of the gravity model that has been proven by such researchers and Anderson and van Wincoop.\(^{53}\) Therefore, the model limits the number of distance variables to the strict necessities. Where possible, it uses continuous variables of distance and proximity; where sensible, it uses logarithmic transformations.

Bilateral trade flows—a sum of the exports recorded by each country in a country pair to the other, smoothed across anomalous spikes and troughs—come from the Correlates of War Project’s International Trade v4.0 dataset, and are recorded in millions of current U.S. dollars (USD).\(^{54}\) These are not transformed to constant USD because the gravity model relies upon assumptions that do not necessarily hold when values are inflation-adjusted. These bilateral trade flows are then transformed by a natural logarithm. This dependent variable is coded as \(\ln(\text{Trade})\).

The variables critical to this study are the distance variables. For physical trade barriers, the model includes the logarithm of great circle distance, in kilometers, between two countries’ average centers of population (coded as \(\ln(\text{Geo Distance})\)), as well as a necessary binary indicator variable, \(\text{border}\), for border contiguity. These data come from the exhaustive dataset of dyadic gravity variables collected by Centre d’Études Prospectives et d’Informations Internationales (CEPII).\(^{55}\)

The demographic distance variables draw on best practices and data from multiple sources. The model includes Helble’s measure of political regimes that combines a general recognition of the effect of democracy on trade with the separate but related effect of the different regime types on trade.\(^{56}\) The variable is the logarithm of the product of each country’s average Freedom in the World score, as constructed based on data from Freedom House.\(^{57}\) The product ranges from one (both countries are highly democratic) to forty-nine (both countries are highly authoritarian), with products in the middle range representing either intensely different regime types or two regimes with equally average political and civil rights. This score is
transformed by the natural logarithm; it is more sensible to interpret the effect on trade of percentage changes in shared authoritarianism than of level changes of an artificial joint authoritarianism score. The variable is:

\[ \ln(\text{Political Score})_{ij} = \ln \ln \left( F_{it} F_{jt} \right) \]

Where \( F_{it} \) and \( F_{jt} \) are the average Freedom in the World scores in year \( t \) for countries \( i \) and \( j \), respectively.

The measures of cultural distance and proximity used in this study include religion, a useful macroscopic indicator of cultural similarity; native language, a key granular measure of cultural distinctness; and spoken language, which falls between the two. The study omitted a direct measure of ethnicity because the data were unreliable and inconsistent across countries and the information is adequately included in the model by native language. It also omitted official language, since official languages are chosen for three reasons already accounted for by the model: it is commonly spoken in a country; it is a relic of imperial hegemony; English is a global lingua franca.

I borrow Helble’s method for measuring religion, which avoids sweeping generalities about countries’ shared affiliations inherent to binary variables, and use the World Religion Project’s National Religion Project v1.1 dataset. For each country, the proportions of the religions in each country with a following greater than or equal to 3 percent of the population are normalized such that these proportions add to one; for each country pair, the normalized population proportions of each of these significant religious groups are multiplied (yielding the probability that two persons, one randomly selected from each country, will share that religion), and sum across all such products for a country-pair. This sum represents the probability that two persons, one randomly selected from each country, would share a religion (it would be zero if nobody shared the religion, and 1 if everyone did). The logarithm of this number enters the model as a continuous-variable measure of religious similarity, which should have a positive coefficient representing the percentage increase in bilateral trade correlated with a one-percentage-point increase in religious similarity. Mathematically, this measure is:

\[ \text{Common Religion}_{ijt} = \sum_{a=1}^{m} r_{ait} r_{bjt} \]

Where: \( i = \) country \( i \) in a country pair; \( j = \) country \( j \) in a country pair; \( r_{ait} = \) the normalized proportion of country \( i \) classified as religion \( a \) at time \( t \); \( r_{bjt} = \) the normalized proportion of country \( j \) classified as religion \( a \) at time \( t \).

The final distance category—linguistic distance—consists of three proximity variables, using Melitz and Toubal’s measures of linguistic proximity. Included are two separate variables indicating the propensity of individuals in countries \( i \) and \( j \) to share a spoken or native language. Melitz and Toubal constructed these in nearly the same manner (and for the same reason vis-à-vis binary variable measures) as the method described above for religion, with slight adjustments. They are coded as Common Spoken Language and Common Native Language, respectively.

Also included is Melitz and Toubal’s rather more granular measure of the linguistic similarity of distinct languages. This linguistic proximity score is based on Levenshtein distance, which measures the number of subtractions, additions, or substitutions needed to turn one string into another. This score is calculated for words with identical meanings from a linguist-determined list of the 100-200 most important words in any given language and adjusted for noise from expected random similarities. The mathematical calculation for this measure is roughly:

\[ DL_{ijt} = \frac{2}{\sum_{a=1}^{2} \sum_{b=1}^{2} dl_{ai,bj} d_{ai,bj}} \]

Where: \( i = \) country \( i \) in a country pair; \( j = \) country \( j \) in a country pair; \( dl_{ai,bj} = \) the reciprocal of an adjusted Levenshtein distance between languages \( a \) and \( b \); \( d_{ai,bj} = \) the normalized proportion of country \( i \) that natively speaks language \( a \) at time \( t \); \( DL_{ij} = \) the weighted average native linguistic distance between countries \( i \) and \( j \) at time \( t \).

Toubal and Melitz transform the measure a few more times for a final measure of linguistic proximity, coded in this study as Linguistic Proximity. Crucially, Linguistic Proximity has a value of zero when the top-two native language are comparable through Common Spoken Language and Common Native Language, lending Linguistic Proximity the interpretation of the effect on trade of native linguistic similarity less directly shared language. Linguistic Proximity, Common
Spoken Language, and Common Native Language are all, unfortunately, time-invariant, but language, in any case, changes little over only thirty-five years. For a more complete discussion, please see Section 3 of Melitz and Toubal.

Key control variables, also from Head’s dataset via CEPII, are dyadic: common currency; common membership in some sort of economic union (this sample contains no examples of currency or customs unions and only some regional trade agreements); common membership in the World Trade Organization or its predecessor, the General Agreement on Tariffs and Trade; a common legal system origin, such as English common law or Islamic law; and a common colonial relationship (country i colonized or shared a colonizer with country j for some significant period).

More information detailing the collection, construction, merger, and cleaning of these data can be found in Appendix A.

MISSING DATA
As tables B1 and B2 in Appendix B clearly show, missing data was a significant, though by no means insurmountable, problem for this study. 76 percent of observations were complete; the remaining 24 percent observations were missing a value for between one and four of the following variables: ln(Trade), Common Spoken Language, Common Native Language, Linguistic Proximity. Very rarely were more than two variables’ values simultaneously missing. Listwise deletion therefore wasted valuable information. Moreover, it was highly unlikely these variables are missing completely at random (MCAR, the only case in which ignoring partially missing data does not bias results) because the missing observations tended to occur in the less developed and integrated countries of the world system, such as Myanmar and Mongolia. Listwise deletion therefore not only would have wasted information, but also biased results. It was also unlikely that missing values were missing at random (MNAR, in which case the value of each missing variable is the only predictor of the same variable’s missingness), and thus it was theoretically possible to meaningfully predict missing values from other variables that may or may not be in the dataset.

The most reasonable method of imputing these missing values was multiple imputation. When complete observations are similar on most measurable variables to a partially missing observation, the missing value is better filled by looking at those similar, complete observations than other observations. Therefore, a linear regression model or other statistical method run on the predictor variables (other than the main method) can better predict the partially missing value than using simple mean replacement or even hotdeck replacement. Because the data exhibited multiple overlapping patterns of missing data, the multiple imputation chained equations (MICE) technique was used to fill in missing values. For more information on this reasoning and methodology, please see Appendix B.

Descriptive statistics for the data with missing values replaced by average imputations are presented in Tables 1 and 2 on pages 22 and 23.

It is evident that common currency, with a mean almost equal to zero and 99th percentile at zero, has no meaningful variation; indeed, in preliminary versions of the study, it confounded significantly with the constant term in the regressions, and was thus removed from further consideration. All other variables in the model varied enough to meaningfully identify their effects and warranted inclusion. Many variables, particularly the indicators, were quite skewed, and although the dependent variable—ln(Trade)—had kurtosis closest to a value of three among the variables in the model, the kurtosis values were often quite distant from the normal distribution’s kurtosis value of three. There were no extraordinarily odd distributions, but the data’s non-Gaussian nature, combined with the model’s usage of fixed effects by country-year and imputed values, called for the usage of robust standard errors.

The strongest correlations were, as expected, among demographic proximity variables, although the exact nature of these correlations, such as the negative relationship between Linguistic Proximity and the other demographic variables, was at first puzzling. In this particular case, as Melitz and Toubal point out in their explanation of these variables, Linguistic Proximity is uninterpretable in the absence of Common Native Language since it represents the linguistic proximity between countries’ non-shared native language populations. Another important correlation to note is the strength of the relationship between ln(Geo Distance) and demographic proximity variables. Other correlations of note were between time and ln(Trade), ln(Political Score), WTO, and Trade Treaty, representing the fact that even this study’s non-Western sample experienced significant globalization and liberalization; the weakness of correlation between time and
In(Geo Distance), border, Common Spoken Language, Common Native Language, Linguistic Proximity, and Common Religion; and between Common Legal System and Colonial Relationship on the one hand and between Common Legal System and the demographic variables on the other. This last fact weakened the regression when Common Legal System and Colonial Relationship were both included; since Common Legal System adds little unique value that is not addressed by either the other cultural variables or colonial status, and estimates using Common Legal System or Colonial Relationship status were interchangeable, Common Legal System was excluded from the published estimates. However, regression estimates including it are available upon request.

Summary statistics and correlations for the original data, pre-imputation, never differed drastically from the post-imputation data. They are included in Appendix B.

Table 1: Imputation Replacement Data Summary Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Min</th>
<th>1st Qu</th>
<th>Median</th>
<th>3rd Qu</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln(Geo Distance)</td>
<td>-3.98</td>
<td>-0.37</td>
<td>0.00</td>
<td>0.30</td>
<td>2.18</td>
</tr>
<tr>
<td>border</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Common Native Language</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Common Religion</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Common Language</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>WTO</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Trade Treaty</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Common Legal System</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Colonial Relationship</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>N</td>
<td>2285</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

METHODOLOGY

The econometric model uses ordinary least squares (OLS) with a regression discontinuity design (RDD) to ascertain the presence of network effects in cultural barriers to trade. This method allows the regression to separate the relationship of demographic and physical distance variables to trade before and after an event that improves intercultural business and economic ties in a subset of the trading cultural blocs. The key, then, is finding an event that improves intercultural business ties directly for only a subset of the cultural blocs in question.

For the sake of robustness, the study used two events in different parts of the data’s chronological scope and with different methods of directly improving intercultural economic ties to ward off confoundedness with other time-related factors. The first discontinuity corresponds to the establishment of diplomatic relations between China and Saudi Arabia in 1990, the second to Turkey’s flotation of the lira’s foreign exchange rate in 2001. These are useful because they represent a significant improvement in a subset of Middle Eastern countries’ ease-of-doing-business with...
East Asia and/or the world.

Diplomatic ties, as shown by Rose, significantly improve international trade. Since the exchange of ambassadors in 1990 between China and Saudi Arabia is the only major change in diplomatic relations between East Asia and the Middle East in the timeframe of this study, it is a superb discontinuity candidate. Meanwhile, the lira’s flotation, as the only major and unilateral change in currency policy in the scope of this study, is a useful discontinuity on account of the smoother business operations that come with free foreign exchange markets. Unlike fixed exchange rates, floating rates do not artificially create foreign currency shortages and surpluses that discourage trade and investment in the proper sectors. Moreover, especially in Turkey’s case, a fixed exchange rate foists dual responsibility over monetary and foreign exchange rates onto the central bank, signaling extensive government economic intervention (and probably mismanagement). Eliminating these obstacles by floating the exchange rate therefore tends to integrate economies despite the associated exchange-rate risk.

It should also be noted that although the flotation of the lira had significant ramifications for Turkey’s balance of trade, this aspect of exchange rate flotation would not affect the bilateral trade flow (a sum of both imports and exports) between Turkey and any other country.

Furthermore, these two events occur roughly one and two-thirds of the way through the chronological range of the data, providing enough data on both sides of these discontinuities for adequate predictive power.

A variable indicating whether an observation represents a bilateral trade flow prior to or after the discontinuity is interacted with the range of distance variables indicated previously. This allows separate identification of the impact of various geographic, political, and demographic trade barriers on trade between East Asia and the Middle Eastern region both before a limited direct improvement in trade relations and after that improvement gives businesses in both cultural regions time to learn from their transactions, diminishing the informal costs associated with intercultural international trade. This significant improvement in intercultural economic relations is expected to diminish the importance of information and trust-based trade costs, and thus the coefficient on the demographic interaction terms should hold the opposite sign of their respective plain demographic variable’s coefficient, which means that the magnitude of the overall coefficient on the demographic proximity variable diminishes. Meanwhile, the interaction on the physical interaction terms should be insignificant.

It is expected that the effect of this discontinuity will take time to appear because international business operates in cycles and contracts limit short-term responses to most international events. There are therefore three econometric specifications for each RDD: the first includes a dummy variable equal to unity in the year of the event’s occurrence, and zero otherwise; the second includes a dummy variable equal to unity one year after the event, and zero otherwise; the third includes a dummy variable equal to unity two years after the event, and zero otherwise. These three specifications of each RDD provide additional robustness and insight into the duration of the international economic adjustment period.

To control for the multilateral resistance factors that are key to a theoretically sound gravity model, a time-varying fixed effect is included for each partner in a country pair (but not for the pair itself, as that would sweep away all variables associated with a country pair—namely everything in this model). This helpful eliminates effects on trade deriving simply from changes in economic circumstances from year to year (such as long-run economic growth, or recession in a few countries), controls for the role of resource extraction in countries belonging to the Organization of Petroleum Exporting Countries (OPEC), and renders the inclusion of other monadic variables, such as a population’s education level or size, or a country’s trade infrastructure, GDP, or status as an island unnecessary. In total, 1,260 fixed effects were absorbed, with two “on” for any given observation (e.g. Saudi1990 and China1990).

The regression equation therefore looked as follows:

$$\ln(\text{Trade}_{ij}) = \beta_0 + \beta_1 \text{DISCONTINUITY}_{i} + \beta_2 \ln(\text{Geo Distance}_{ij}) + \beta_3 \text{DISCONTINUITY}_{j} + \beta_4 \ln(\text{Political Score}_{ij}) + \beta_5 \ln(\text{Language}_{ij}) + \beta_6 \ln(\text{Religion}_{ij}) + \beta_7 \ln(\text{Culture}_{ij}) + \beta_8 \ln(\text{Trade Tutorial}_{ij}) + \epsilon_{ij}$$
RESULTS

Table 3 below compares predictions for the sign of each variable with the actual sign results, while Tables 4 and 5 on the following pages display the full results of the main regressions.

<table>
<thead>
<tr>
<th>Table 3: Parameter Predictions and Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>constant</td>
</tr>
<tr>
<td>DISCONTINUITY event</td>
</tr>
<tr>
<td>ln(Geo Distance)</td>
</tr>
<tr>
<td>border</td>
</tr>
<tr>
<td>ln(Political Score)</td>
</tr>
<tr>
<td>Common Spoken Language</td>
</tr>
<tr>
<td>Common Native Language</td>
</tr>
<tr>
<td>Linguistic Proximity</td>
</tr>
<tr>
<td>Common Religion</td>
</tr>
<tr>
<td>WTO</td>
</tr>
<tr>
<td>Trade Treaty</td>
</tr>
<tr>
<td>Colonial Relationship</td>
</tr>
</tbody>
</table>

As a first remark, all specifications have high explanatory power, passing an F-test for goodness-of-fit and possessing adjusted $R^2 > 0.5$. This confirms the importance of a cultural gravity model in explaining trade patterns. Secondly, there is extreme consistency in results across lag periods, with the only variations in the sign of Saudi-China compared with Common Native Language and Saudi-China compared with Linguistic Proximity in lags one and two, and the significance levels of a few coefficients, such as between Saudi-China compared with border and Lira compared with border or Saudi-China compared with Common Spoken Language and Lira compared with Common Spoken Language. This lends robustness to the results, messy as they are.

As Table 3 shows, the results broadly confirm theories from the literature. These results confirm that physical distance is a statistically significant drag on trade volumes, with a 1 percent increase in distance diminishing trade by as much as 1.97 percent, and that border contiguity is a significant boon. A common colonial relationship increases trade by around 67 percent.

The key to understanding these odd results lies in a few data curiosities, first regarding the scope. The negative correlation of trade agreements with trade volumes may be explained by the fact that many countries in sample lack fruitful trading partners within the sample, and instead sign trade agreements designed to facilitate supply chains for a more powerful, external country—such as the United States signing trade agreements with Japan, South Korea, and Taiwan. Likewise, religious similarity’s salutary effect on trade would have been underestimated due to the exclusion of economically massive trading partners closer to one cultural bloc than another, such as the European Union.

Political relationships were so insignificantly affected by the discontinuity event that the DISCONTINUITY compared to ln(Political Score) terms were always omitted due to multicollinearity, presumably with the fixed effects and baseline ln(Political Score). Likewise, Lira compared with border was insignificantly different from zero. This fits the hypothesis that non-cultural variables would not be heavily impacted by the intercultural economic event.

As the highlighted sections of Table 3 show, there are also clearly some unpredicted results that require explanation. WTO membership seems to have little impact on trade, while a trade agreement correlates with significantly diminished trade (as much as 16.2 percent). A 1 percent increase in the proportion of two countries’ populations that speak the same language or that follow the same religion correlates with as much as 408 percent and 287 percent less trade, respectively, while an event of intercultural business importance seems to attenuate this affect by as much as 147 percent and 360 percent, respectively. Moreover, the event seems to attenuate the dissuasive power of physical distance on trade by as much as 0.6 percent, more than a third of distance’s original impact.

The results also tentatively confirm this research paper’s hypotheses. Not only are all cultural proximity variables critically significant at baseline, but all their discontinuity interaction terms (with the exceptions of Common Native Language and Linguistic Proximity in lags one and two) carry the opposite sign of the baseline and are jointly significant. This means that although a 1 percent increase in the proportion of populations that natively speak the same language increases trade volumes by as much as 540 percent, an event that brings businesses into contact between a subset of the Middle East and East Asia attenuates this effect, knocking off 75 percent of that impact. This relationship holds across lags, with only slight changes in the precise nature of the relationship, and across both the establishment of Saudi-Chinese relations in 1990 and the flotation of the Turkish lira in 2001.
Union, which is undeniably closer to the Middle East. Such omitted economies, whose sizes are necessarily uncontrolled for since they are excluded from study, deflect trade away from intra-cultural trading partners that seem otherwise poised to be prime trading partners based on cultural proximity and physical distance. This flaw would misattribute the lack of intra-religious trade to the religious similarity of two countries rather than their mutual physical proximity to more advanced economies.

Table 2 points to the third issue. The high correlation of the demographic variables with physical distance makes separate identification difficult. As a result, the DISCONTINUITY compared to ln(Geo Distance) terms are statistically significant and confound with the demographic discontinuity interactions. It also makes some sense that Common Religion, in this study’s scope, is negatively associated with trade volumes. Helble found that Buddhist countries are slightly discouraged from intra-religious trade.56

Combining the variable correlations, the high-dimensional fixed effects, and unique attributes of the limited country scope, it is a wonder that the demographic-discontinuity interaction terms so consistently yielded coefficients carry the opposite sign of their baseline with joint significance.

At the same time, one should not dismiss the fact that in nearly every case—even where the baseline coefficient bucked the theoretical prediction—the demographic-discontinuity interaction terms carried the opposite sign of their corresponding baseline term, with the intercultural business discontinuity thus attenuating whatever effect of cultural proximity previously prevailed. This clearly indicates consistent network effects. Rather than diminishing the power of demographic similarity to encourage trade (or the power of demographic distance to discourage trade), improvement in economic ties between subsets of cultural blocs simply seems to make culture less relevant to trade overall, both between and within each cultural bloc. In other words, demographic barriers move closer to having no effect on trade, no matter where they started.

<table>
<thead>
<tr>
<th>Table 4: Discontinuity for establishment of Saudi-China ties in 1990</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>ln(Trade)</td>
</tr>
<tr>
<td>constant</td>
</tr>
<tr>
<td>Saudi-China Relations Discontinuity</td>
</tr>
<tr>
<td>ln(Geo Distance)</td>
</tr>
<tr>
<td>ln(Geo Distance)</td>
</tr>
<tr>
<td>Saudi-China/ln(Geo Distance)</td>
</tr>
<tr>
<td>ln(Industry)</td>
</tr>
<tr>
<td>ln(Political Score)</td>
</tr>
<tr>
<td>ln(Political Score)</td>
</tr>
<tr>
<td>Common Spoken Language</td>
</tr>
<tr>
<td>ln(Industry)</td>
</tr>
<tr>
<td>ln(Political Score)</td>
</tr>
<tr>
<td>ln(Geo Distance)</td>
</tr>
<tr>
<td>ln(Political Score)</td>
</tr>
<tr>
<td>ln(Geo Distance)</td>
</tr>
<tr>
<td>Common Spoken Language</td>
</tr>
<tr>
<td>ln(Geo Distance)</td>
</tr>
<tr>
<td>ln(Political Score)</td>
</tr>
<tr>
<td>ln(Geo Distance)</td>
</tr>
<tr>
<td>Common Spoken Language</td>
</tr>
<tr>
<td>ln(Geo Distance)</td>
</tr>
<tr>
<td>ln(Political Score)</td>
</tr>
<tr>
<td>ln(Geo Distance)</td>
</tr>
<tr>
<td>Common Spoken Language</td>
</tr>
<tr>
<td>ln(Geo Distance)</td>
</tr>
<tr>
<td>ln(Political Score)</td>
</tr>
<tr>
<td>ln(Geo Distance)</td>
</tr>
</tbody>
</table>

(Robust standard errors in parentheses)

<table>
<thead>
<tr>
<th>x</th>
<th>p &lt; 0.1</th>
<th>p &lt; 0.05</th>
<th>p &lt; 0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>22050</td>
<td>22050</td>
<td>22050</td>
</tr>
<tr>
<td>R-sq</td>
<td>0.897</td>
<td>0.897</td>
<td>0.897</td>
</tr>
<tr>
<td>Adj. R-sq</td>
<td>0.892</td>
<td>0.892</td>
<td>0.892</td>
</tr>
<tr>
<td>F</td>
<td>28.93**</td>
<td>28.93**</td>
<td>28.93**</td>
</tr>
</tbody>
</table>
This study tentatively indicates that there are network effects in trade between cultural blocs such that small investments in intercultural trade overall attenuate the effect of demographic similarity on trade. The effect holds true regardless of whether cultural similarity was originally a boon or drag on trade; the increase in intercultural business puts the two cultural blocs in question on an overall more equal playing field. Moreover, this attenuation in response to the growth of business acumen abroad is far faster than the realities of contractual rigidity and learning curves might lead one to believe. One explanation lies in rational expectations, but a more likely explanation is the “fail fast, then scale up” attitude demonstrated in Barkema and Drogendijk.\textsuperscript{55}

A larger sample—accounting for all relevant dyads and increasing the number of observations per coefficient—is necessary for a more definitive attribution of these network effects to separately predict the effect of intercultural business on cultural trade barriers from its effect on other trade barriers. It may be entirely possible that network effects exist in all barriers.

**IMPLICATIONS**

Trade theorists should reconsider the underlying assumption in gravity theory that actors choose to trade with each other independent of the other trade relationships in their network, since this study presents evidence that directly contradicts that assumption. Trade policymakers, meanwhile, have a new tool to deal with the economic implications of the “clash of civilizations.” Rather than write off intercultural economic relations and focus inward or work to outcompete other cultures, policymakers can prioritize economic relationships with those foreign countries that are central to their cultural bloc’s economy and simultaneously are more inclined to do business with the policymakers’ home country. The returns to such trade will accrue to economic relations with other members of the foreign bloc as well.

This study also provides citizens—particularly in the Western and “Orthodox” worlds—a warning against tribalist demagogues that build their power off fears of civilizational decline. By breaking delicate bridges between high- and low-growth areas, these demagogues build a cascade of cultural barriers that will harm ordinary citizens’ wellbeing.

**CONCLUSION**

This study had numerous flaws that future research...
can and should address. The first category is data-related. Obviously, complete data would eliminate the trade-off of information and bias inherent to missing data analysis. The acquisition of reliable, standardized ethnicity data would add greatly to the granularity of cultural analysis, as would time-variant data on language. Future research could include indicator variables for cultural bloc, as well. Such an addition would be required to expand the RDD approach to a larger scope spanning all dyads in the global trading network, since the discontinuity would be limited to countries in one of the cultural blocs affected by the economic event. This would enable including countries whose civilizational heritage is too complicated to include in a two-bloc analysis, such as Indonesia and Malaysia in this study, as well as other countries outside either civilization impacted by the discontinuity event. As the discussion of unexpected results for the effect on trade of religion, trade agreements, and other variables made clear, a scope that sections off large parts of the world is subject to much error.

Another fruitful avenue of research would be measuring network effects with an entirely different method than RDD. An RDD only estimates the regression across two time periods and presumes an event’s exogenous effects on intercultural business. A future study may try a different approach less vulnerable to confoundedness with other chronological effects, such as interacting physical, political, and demographic distance variables with the sum of the previous five years’ bilateral trade volumes. This would test a different but related question of whether there are increasing returns in intercultural trade to overall intercultural trade, rather than whether the development of conduits through cultural barriers improves overall intercultural trade. Alternatively, the RDD approach could be maintained, but on a global scale.

Although this study lends further credence to Armington’s “home bias” and the recent literature that home bias derives from both physical and cultural distance, it also warns of two problems in the literature around the gravity model of trade. Firstly, case studies excluding large economies should no longer be pursued. Secondly, the gravity model’s assumption that decisions to trade parcels between two economies are independent of all other decisions must be reexamined.

APPENDICES

APPENDIX A

Trade data in the Correlates of War project’s International Trade dataset was only recorded as zero if the importer and the exporter both reported zero trade in both directions in a given year to the IMF and every other source consulted for compiling the data set. If all simply did not report data, trade was labeled missing. I therefore safely assumed zero-trade values to in fact be indistinguishable from zero and replaced these values with 0.000001—that is, $1.00—to permit a logarithmic transformation as the gravity model specifies (cf. leaving zero-trade: \( \ln(0) = \{-\} \)).

Common Religion

Raw data on religious composition of countries across the world came in five-year intervals between 1945 and 2010 as the proportion of the population of a given country following each religious sect or religious grouping in a given year. First, I duplicated each religious proportion value for the next four years to have annual data between 1980 and 2014. Next, I kept data on proportions following Christianity, Judaism, Islam, Buddhism, Hinduism, Zoroastrianism, Taoism, Confucianism, and Shintoism, and set groups with less than 3 percent of a population to 0 percent. Past papers used a 4 percent or 5 percent cutoff to avoid overly granular data that make calculations difficult and prone to outlier data points unrepresentative of religious similarity, but 3 percent better measures differences in national culture owing to minority groups—for example, a 4 percent or 5 percent cutoff would imagine the United States to have never had a Jewish or Muslim presence, despite these minorities’ significant role in modern American culture compared to otherwise similar countries in modern Europe. I merged these proportions to the master dataset by year \( t \) and country \( i \) and by year \( t \) and country \( j \), then multiplied the proportion of country \( i \) following a religion in year \( t \) by the proportion of country \( j \) following the same religion in year \( t \). Summing across religions, for each observation, provides a number roughly representing the chance that a random individual from country \( i \) and a random individual from country \( j \) would share any religion in year \( t \).

There were two issues with this measure. The first is dual religion. Common Religion is not truly a probability of shared religion; it can theoretically be
larger than 1, e.g. everyone in both countries follows
the same two religions. Luckily, dual religiosity was
only a major occurrence in Japan, where many are
Buddhist and Shintoist, and China, where many follow
a syncretic custom in addition to Buddhism or Taoism;
syncretism was excluded from Common Religion on
account of its inconsistency across borders.

A more significant structural issue was Islamic
sectarianism. It became clear that the damage done to
Common Religion by counting differences between
Shi’a, Sunni, and Ibadi populations on a par to those
between Buddhists and Christians would be far greater
than that done by ignoring sectarian differences. Much
modern Islamic sectarianism is, in any case, geopoliti-
cal rather than cultural in nature; ln(Political Score)
and the fixed effects partly control for this.

\[ \text{ln}(\text{Political Score}) \]

The joint authoritarianism value derives from scores
on political rights and civil liberties for each country
and territory in the world (with 1 the freest and 7 the
most authoritarian), 1973-2018, reported by Freedom
House. From 1981-1989, survey data did not corre-
spond to the calendar year; instead, there were eight
samples covering these nine years. The 1981 survey
period extended past 1981 to August 1982; the 1983
survey period extended from August 1982 through
November 1983. The next seven survey periods ran
November-November until a return to survey periods
matching calendar years in 1990. I set these survey
periods to the year of which they covered a major-
ity, which left 1982 blank. I therefore created 1982
observations for all countries by averaging the 1981
and 1983 scores for political rights and civil liberties,
respectively.

After reshaping the data from lists of political
and civil scores by year for each country into a list of
country-year observations with a political and civil
variable and merging this data into the dyadic dataset,
I averaged the political rights and civil liberties scores
to create composite “freedom indices” for country i in
year t and country j in year t. Multiplying these two
freedom indices for each member country of a dyad
yielded a joint freedom score.

APPENDIX B

Rubin (1976), Little and Rubin (2002), Schafer &
provide justification for using multiple imputation to
avoid losing information or biasing results. The data’s
missingness patterns are described in tables B1 and
B2.

<table>
<thead>
<tr>
<th>Table B1: Missing Data Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>ln(Trade)</td>
</tr>
<tr>
<td>Common Spoken Language</td>
</tr>
<tr>
<td>Common Native Language</td>
</tr>
<tr>
<td>Linguistic Proximity</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table B2: Missing-value Patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pattern (1 means complete, 0 means missing)</td>
</tr>
<tr>
<td>Percent</td>
</tr>
<tr>
<td>76 percent</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>100 percent</td>
</tr>
</tbody>
</table>

I imputed 100 times to get a stable average imputation
for each missing value, using (in addition to ln(Geo
Distance), border, ln(Political Score), Common Reli-
gion, WTO, Trade Treaty, Common Legal System,
and Colonial Relationship, the key variables from the
model) GDP per capita for ln(Trade) (and any avail-
able data on GDP per capita for that country in other
pairings and years to predict the occasional missing
GDP per capita value) and a dummy of common eth-
ic language spoken by at least 9 percent of the popu-
lation of both countries in a dyad for Common Spoken
Language, Common Native Language, and Linguistic
Proximity; both of these variables are from CEPII’s
gravity dataset. I chose to use the regression multiple
imputation method rather than simple variable mean
or hotdeck replacement because multiple imputation
better fits missing values into the spectrum of the data;
meanwhile, I did not use logit imputation even to
predict the probabilistic *Common Spoken Language*, *Common Native Language*, and *Linguistic Proximity* because they are continuous dependent variables (in the context of the imputation model).

The established *mi estimate* command to combine the main model’s specified regressions, run on each complete set did not work with high-dimensional fixed effects; meanwhile, simply running the main regressions on all imputation-completed datasets together would bias standard errors downward by pretending the dataset contained 2,205,000 observations instead of 22,050. The average of the one-hundred imputed values for each case where the original value was missing was therefore calculated and these average imputations were used as replacements for the value of the missing observations.

Some of the original one-hundred imputed values of *Common Spoken Language*, *Common Native Language*, and *Linguistic Proximity* were negative because the imputation model was linear, without a minimum possible imputation. However, less than 5 percent of imputations were problematic in this way, and, after replacing missing values with average imputations, no observations’ values presented this problem; it was therefore unnecessary to make any case-by-case adjustments. Similarly, some imputations of ln(Trade) were arbitrarily low (e.g. -20, indicating bilateral trade at $0.000000001 million = $0.001), but this was not a problem because, like the $1 stand-in for zero-trade, this imputation simply represents trade values equivalent to zero.

The pre-imputation descriptive statistics are presented in Tables B3 and B4.

<table>
<thead>
<tr>
<th>Table B3: Original Data Summary Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>ln(Trade)</td>
</tr>
<tr>
<td>ln(Telecom)</td>
</tr>
<tr>
<td>ln(Boxer)</td>
</tr>
<tr>
<td>ln(Political Score)</td>
</tr>
<tr>
<td>ln(Trade)</td>
</tr>
<tr>
<td>ln(Telecom)</td>
</tr>
<tr>
<td>ln(Boxer)</td>
</tr>
<tr>
<td>ln(Political Score)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table B4: Original Data Correlations Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>(correlations among complete observations)</td>
</tr>
<tr>
<td>Year</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>0.374</td>
</tr>
<tr>
<td>0.257</td>
</tr>
<tr>
<td>0.367</td>
</tr>
<tr>
<td>0.652</td>
</tr>
<tr>
<td>0.226</td>
</tr>
<tr>
<td>0.257</td>
</tr>
<tr>
<td>0.652</td>
</tr>
</tbody>
</table>

Endnotes

3. Pinelopi K. Goldberg, Development economics
5 Ibid.
21 Paul S. Armington, “A Theory of Demand for Products Distinguished by Place of Production (Une Théorie de La Demande de Produits Différenciés d’après Leur Origine) (Una Teoría de La Demanda de Productos Distinguidos Según El Lugar de Producción),” *Staff Papers (International Monetary Fund)* 16, no. 1 (1969): 159–78.
467–517.
32 Melitz, “Language and Foreign Trade.”
33 Melitz and Toubal, “Native language.”
36 Helble, “Is God Good for Trade?”
40 Ibid.
41 Gokmen, “The Impact of Cultural Differences on Trade.”
48 Dunlevy and Hutchinson, “The Impact of Immigration.”; Aker et al., “Borders, Ethnicity, and Trade.”
50 Savage and Deutsch, “A Statistical Model.”
51 Huntington, The Clash of Civilizations.
53 Anderson and van Wincoop, “Gravity with Gravitas.”
56 Helble, “Is God Good for Trade?”
58 Zeev Maoz and Errol A. Henderson, “The
59 Melitz and Toubal, “Native Language.”
60 Ibid.
61 Ibid.
62 Rose, “The Foreign Service and Foreign Trade.”
64 Helble, “Is God Good for Trade?”
65 Barkema and Drogendijk, “Internationalising in Small, Incremental or Larger Steps?”
66 Armington, “A Theory of Demand.”
67 Savage and Deutsch, “A Statistical Model.”

Bibliography


Bougheas, Spiros, Panicos O. Demetriades, and Edgar L.W. Morgenroth. “Infrastructure, Transport


Kim, Sangmoon, and Eui-Hang Shin. “A


TUNISIA’S ENNAHDA PARTY: 
DEVELOPING A FRAMEWORK FOR INTERPRETING POLITICAL DECISION-MAKING IN HISTORICAL CONTEXT

STEPHEN GARRETT, GEORGETOWN UNIVERSITY CLASS OF 2019

ABSTRACT

Tunisia’s post-revolution democratic transition is the subject of a large literature and draws attention for its relative success and its ongoing debates and challenges. This thesis analyzes the Tunisian political party Ennahda, particularly its history and behavior, in order to develop a framework for understanding the contextual basis for contemporary policies and actions. In particular, the thesis answers the following questions: What does Ennahda seek to achieve as a party? How is Ennahda impeded from achieving its goals? How is Ennahda acting to overcome challenges in an effort to reach its desired ends? The thesis first determines the party’s goals by analyzing party media, election data, interviews with party members, and a variety of other sources. Following the explication of goals, the thesis discusses the roots of present-day opposition to Ennahda by exploring the party’s historical relationship to the Tunisian political establishment, regional political movements, and 2011 revolution. Finally, the thesis explores a method for understanding recent party actions as comprising a strategy to achieve goals in light of opposition. It does so by analyzing a number of party decisions and policies as well as party rhetoric. The thesis concludes that Ennahda is taking action by emphasizing its Tunisian identity, moving away from political Islam, and adopting moderate political strategies in an ongoing effort to overcome obstacles and achieve party goals.

ACKNOWLEDGEMENTS

Thank you to my advisor, Yvonne Haddad, for pushing me to write everything down and make room in my head for new ideas. Thank you to Tamara Sonn for introducing me to Tunisia and Ennahda. Thank you to Dr. John Esposito, Dr. Sarah Yerkes, Dr. Noureddine Jebnoun, and Dr. Radwan Masmoudi for providing insight in the early stages of the research process. A particular thank you to Dean Lisa Gordinier for her invaluable and compassionate support throughout this experience and to Dean Emily Zenick for holding me accountable and offering sage advice and encouragement. Thank you to all of my Tunisian friends and mentors, who inspired this thesis. Finally, I would like to thank my parents for their guidance, encouragement, and love through this process and always. This thesis is dedicated to my Nana, Rita Holland.

On December 17, 2010, a young man named Mohamed Bouazizi lit himself on fire in the heart of Tunisia, a small country in North Africa that sits at the confluence of Algeria to the west, Libya to the east, and the Mediterranean Sea as well as Europe to the north. Bouazizi was protesting against harassment by the Tunisian police who worked for a repressive political and social system led by President Zine El Abidine Ben Ali. By the time Bouazizi died of his wounds on January 4 just eighteen days later, popular protests sparked by his self-immolation had already spread throughout the country, including to the capital Tunis. Fueled by corruption, Ben Ali’s repressive regime, rising food prices, and unemployment, protestors used social media to organize massive demonstrations against the President, who had ruled...
Tunisia since 1987. On January 14, 2011, Ben Ali boarded a helicopter outside his coastal palace and fled to Saudi Arabia. The demonstrations that led to Ben Ali’s departure, sparked by Bouazizi’s self-immolation, led to further protests across the Arab world—a phenomenon that came to be known as the Arab Spring. As the world turned its attention toward the ongoing protests in other countries, a new era began in Tunisia.

On January 30, a plane departed from London and landed in Tunis carrying Rached Ghannouchi. Greeted by thousands, Ghannouchi, the leader of the Ennahda Party, stepped into his country for the first time since May 1989. The return of Ghannouchi and his Ennahda party, which he first co-founded as the Mouvement de la Tendance Islamique (MTI) or Islamic Tendency Movement in 1981, represented a new moment for the country. Ennahda was a party that promoted political Islam or Islamism (the terms are interchangeable for the purposes of this paper), meaning it promoted the notion that the religion of Islam “provides guidance for all areas of human life, individual and social,” and therefore should guide political institutions and governance. Ben Ali had banned the party from political participation in 1989, led an extended campaign of arrests, and excluded members of the party from society. With Ben Ali’s deposition in 2011, however, Ghannouchi and Ennahda returned to a Tunisia filled with hope for a new era. Almost a decade later, the country has undergone momentous shifts, but questions about the future of the country still remain. The revolution in early 2011 led to a series of popular revolutions around the Arab world known as the Arab Spring; of those revolutions, only Tunisia’s led to a democracy that stands today. The country of about twelve million is not yet politically stable, but it is certainly politically vibrant. Ennahda and other parties contend for popular support as the political system continues to establish norms and patterns, and as society shapes conceptions of Tunisia’s identity.

Ennahda is a particularly compelling party to examine because, like Tunisia, it has undergone tremendous change since the revolution. Unlike most of the political parties that emerged after the 2011 revolution, Ennahda has a long history. Transitioning from an underground movement to an influential political party required a shift in strategy and organisation. The party and its leadership have made noteworthy compromises in the interest of establishing a stable, sustainable democracy in Tunisia, but it has not been a matter of pure magnanimity. Ghannouchi contends that the party, in its own best interest, has worked “for compromise and reconciliation rather than exclusion or revenge.”

This thesis develops a novel framework for understanding the contemporary actions of Tunisia’s Ennahda Party in a historical and political context. The framework called the goals-obstacles-steps framework, will be used to uncover possible motivations for Ennahda’s actions. Simply put, Ennahda’s actions can be categorized as fitting into one or more steps, which are efforts to overcome historically-rooted obstacles to the party’s goals.

Negative public perception of Ennahda, Tunisia’s largest Islamic political party, is rooted in the party’s historical role as an opposition movement to the Bourguiba and Ben Ali regimes, its association with political Islam in the region, and its actions after the 2011 revolution. In an effort to allay public opposition and achieve party objectives such as maintaining national stability, broadening its voter base to win elections, and increasing western support for Tunisia’s democratic transition, Ennahda is working to emphasize its Tunisian roots, move away from political Islam, and adopt moderate political strategies.

This thesis is based on qualitative research of both qualitative and quantitative data. The thesis will continue with a review of existing scholarship and literature pertinent to its development, an explanation of the methodology used to develop this framework, and the data section. The last section is divided into three chapters. Chapter one defines and examines Ennahda’s current goals and how they have evolved over time. Chapter two defines and examines the obstacles that have inspired opposition to Ennahda and thus inhibit the party from achieving its goals. Chapter three defines and examines the steps that Ennahda is undertaking to overcome its obstacles and achieve its goals. The order of these chapters is not chronological, rather it is such that the reader can understand the framework in a logical order. Knowing the obstacles after the goals allows the reader to more readily understand the significance of the barriers in place. Concluding with steps allows the reader to fully understand the existing context for Ennahda’s actions.

LITERATURE REVIEW
The body of literature pertaining to the themes of this thesis is enormous and filled with extensive schol-
arship and analysis. I have examined a range of literature to account for diverse perspectives. For the purposes of developing a systematic review of the available research, the following are the main themes of relevant literature: analyses of Modern Political Movements/Parties; Tunisia’s history; 2011 Revolution; Ennahda’s history; Ennahda’s ideology; Sheikh Rached Ghannouchi; and recent Tunisian politics. Many pieces address more than one of these themes. For the purpose of brevity only some pieces are listed.

ANALYSES OF MODERN POLITICAL MOVEMENTS/PARTIES
To define the position of this thesis and its framework in the context of broader literature, it was necessary to develop an understanding of literature on international political movements. There is an impressive range of academic studies and analyses in this area. Many analyses take a broad view by generally assessing political movements and parties. In this vein, the most important work for the purposes of this thesis is David Easton’s “An Approach to the Analysis of Political Systems,” which offers a model for understanding how a political system functions. A political party is a political system by Easton’s definition. Easton’s model is defined as the inputs – political system or processes – outputs model. The goals-obstacles-steps framework discussed in this paper is different in its distinction between goals and obstacles as competing factors that influence actions, or steps. In addition, the notion of an obstacle category offers a compelling answer to the question of why not every political party achieves its goals in due time. The inputs – political system or processes – outputs model offers “feedback” as a possible constraint on political party action, but this is a response and not a pre-emptive hindrance. The goals-obstacles-steps framework explicitly allows for constraints that are rooted in a different chronological period from the subject of analysis i.e. constraints on past party action that also inhibit action in the period of analysis. In addition to Easton’s work, a number of other political analyses were useful, including those of Germany’s Christian Democratic Union (CDU), Turkey’s Justice and Development Party’s (AKP) policy toward the Kurds, Egypt’s Muslim Brotherhood, and South Africa’s ANC. These and others were useful in building awareness of existing methodologies, particularly with regard to political environments with which I lacked familiarity. While this body of literature helped me to form the framework, the framework is only useful insofar as it is supported with a careful attuning to its subject. In the case of Ennahda, forming a deep knowledge of the country’s political history, as well as the history of Ennahda itself, was necessary.

TUNISIA’S HISTORY FROM 1956
In approaching the subject of Tunisia’s Ennahda party and the broader topic of Tunisian politics in a post-revolution setting, it is necessary to provide an overview of Tunisia’s history. While Tunisia’s history stretches back millennia, the focus of this paper requires analysis primarily of post-colonial Tunisia i.e. after 1956, the year the country achieved independence from France. Scholarship on the subject is often thematic; Tunisia is a small country but can be richly analysed through a host of conceptual lenses including nationalism, political Islam, and modernisation. Some of the most compelling scholarship on Tunisia’s history approaches the country by focusing on one or more of these concepts including Tunisia’s modernization under Habib Bourguiba, the development of its one-party political system, state-building and development, nationalism’s growth from France’s foray through the 1920s, etc. Kenneth Perkins’ History of Modern Tunisia looks holistically at the country as a crossroads between east and west and at its societal and political discourse incorporating both elements to define its unique character. A subset of scholarship on modern Tunisian history deals with the role of Islam, particularly how it functioned politically under Bourguiba and Ben Ali. These sources construct a picture of the debates that underlie Tunisian society and politics to this day. Specifically, they address conflict over what it means to be Tunisian. The roles of colonial memory, differing views of religiosity, and even varying language ability all affected the debate. After 1956, Tunisians worked to interpret differing notions of the Tunisian identity. Understanding the unique history of Tunisia and these debates, particularly with respect to their effect on the country’s contemporary political discourse, is fundamental in constructing an argument about the nature of Tunisian political change.

POLITICAL ISLAM
The discourse on political Islam and its accurate terminological usage is important to understand. The concept of political Islam ties the role of Ennahda in Tunisian politics to a broader academic narrative of interaction between politics and religion, particularly
Islam. The concept also serves as a reference point for analyzing the party. While articles and books on political Islam are numerous, and the arguments are manifold, there are two areas of focus relevant to this thesis. The first is political Islam as it pertains to Islamic political parties now and historically (i.e. beginning with post-colonial context). The second deals with political Islam more broadly in its contemporary usage in scholarship. It is particularly useful to research the difficulty mainstream Islamic political parties have defining their Islamic identity, especially in a world where radical extremists who claim to represent Islam have influenced popular definitions of political Islam.

**ENNAHDA’S HISTORY**

Scholarship on Ennahda’s roots varies in its specificity and scope. Anne Wolf’s 2017 book *Political Islam in Tunisia* is quite useful; it offers a detailed look at Ennahda’s place in Tunisian society from its earliest days until 2016. Rory McCarthy published a similar book in 2018 called *Inside Tunisia’s al-Nadha*. While Wolf and McCarthy focus on Ennahda as a political grouping, including the intra-group dynamics, relationship to other Tunisian parties and politicians, etc., Mohamed Elchami Hamdi (1998) does an excellent job analyzing Ennahda’s role as an Islamic party, focusing on its ideological characteristics and how its changing ideology affected its relationship with Tunisian authorities. Cavatorte and Merone (2013) argue that it is not the party’s changing strategy or group dynamics that most affected Ennahda’s history but rather Bourguiba and Ben Ali’s exclusion of the movement that transformed Ennahda by moving it from a “fundamentalist to conservative party.” Dunn agrees that exclusion was the cause of ideological changes but does so with different implications, believing that exclusion by Bourguiba fomented a temporarily revolutionary ideology in the party that eventually became conservative. Dunn and Cavatorte have somewhat incompatible arguments, as combined they would suggest that a single independent variable- exclusion by the governing authority- led to multiple dependent variables, namely moving Ennahda from conservatism to revolutionary ideology and moving Ennahda from revolutionary ideology back to conservatism. This thesis examines the impact of exclusion on Ennahda’s trajectory before the revolution but puts greater focus on the current effects of exclusion on perception of Ennahda. MTI/Ennahda is less than fifty years old, and its history is still very much alive in the memories of some of its most important members.

**ENNAHDA’S IDEOLOGY**

Coupled with Ennahda’s history is its ideology and agenda as a movement and consequently as it transitioned into a party. On this subject there a multiple sources such as external analyses, party statements and pamphlets, essays and speeches by party leader Rached Ghannouchi, and more. In addition, a piece from the Carnegie Endowment for International Peace analyses the economic agenda of Ennahda, including fiscal, tax, and investment policy, unemployment and poverty, governance, and trade and integration. In a 2016 Foreign Policy article, Karina Piser addresses Ennahda’s ideology and post-revolution goal of maintaining Tunisian democracy. One major shift in recent years, and an important part of this thesis, is Ennahda’s move away from political Islam and Islamism toward Muslim democracy. This move, taken explicitly in 2016, was a change in the party’s self-identification to a new concept which highlighted Ennahda’s support for Tunisia’s democracy. This change indicated a symbolic break from regional Islamic political movements, as well as a terminological disassociation with loosely defined “political Islam” and “Islamism.” Chapter 3b discusses this topic further. An interesting moment in the field of Tunisia studies occurred when in August 2015 Monica Marks published a piece for the Brookings Institution’s *Rethinking Political Islam* series titled “Tunisia’s Ennahda: Rethinking Islamism in the Context of ISIS and the Egyptian Coup” that discussed Ennahda’s response to regional extremism and its methods of reframing its identity. Specifically, Marks narrows in on how regional and local political developments (e.g. the rise of Salafi-jihadism and the 2013 coup in Egypt) affected the way in which Ennahda presents itself as Islamic. Marks analyzes how Ennahda’s move to be a Muslim Democrat party was inspired by the Justice and Development Party in Turkey and the Christian Democratic Union in Germany. She discusses the complaints that many members of Ennahda have about the Egyptian Muslim Brotherhood (MB), a prominent regional Islamist movement (see Chapter 2b). These complaints center on the MB’s focus on its short-term goals that not only prevent the MB from achieving long-term success but also prevent external groups like Ennahda from convincing opponents that Islam and democracy are compatible. Marks also goes into Ennahda’s approach to Salafism.
The following March, Sayida Ounissi Member of Parliament of Ennahda responded to the piece, supplementing and clarifying it. Marks herself later responded to Ounissi’s piece, further emphasizing the roots of Ennahda’s ideology and its methods of identity revision. Both pieces dive into Ennahda’s Tunisian identity, its brand of Islamic democracy, and its inclusive attitude towards its primary partisan opponent, Nidaa Tounes (see Chapter 1b). Marks, in her first piece, focuses specifically on the deleterious effects of external events related to political Islam on Ennahda’s reputation, and the party’s response. This thesis also focuses on Ennahda’s attempt to alter its reputation, but does so with a more historical scope, attributing much of Ennahda’s current reputation to its vilification under Bourguiba and Ben Ali. It also broadens the scope of Ennahda’s goals and more deeply explains relevant concepts.

SHEIKH RACHED AL-GHANNOUCHI

The Ennahda party’s ideology is also intrinsically linked to its co-founder and current leader, Sheikh Rached al-Ghannouchi. Since the 1960s, Ghannouchi has been fairly transparent with his ideas, giving speeches and writing about Islam and politics. The most important and detailed book on Ghannouchi is Azzam Tamimi’s 2001 book, Rachid Ghannouchi: A Democrat within Islamism, which provides a look at the man and his ideological evolution. Tamimi worked closely with Ghannouchi while compiling this book which includes extended summaries of some of Ghannouchi’s untranslated works. Interviews with Ghannouchi are also useful. These interviews and writings also show the way Ghannouchi’s language has changed, becoming more measured around issues of Islam and politics over time. For example, in Linda Jones’ (1988) translation of Ghannouchi’s 1984 article on the deficiencies of the Islamist movement, Ghannouchi uses terms such as Islamism and shari’a that in later writing he stopped using, given their controversial connotation. Jones also has a piece from 1988 that serves as both an analysis of Ghannouchi’s thought and also as a history of Ennahda (and its previous name, MTI) up to that point. In that piece, Jones is optimistic about Ennahda finding a place in Tunisia’s political scene after Ben Ali’s rise to power, later proven wrong when Ben Ali closed the political system. As its leader and co-founder, Ghannouchi’s ideas serve as a lens through which to interpret party actions. As Ennahda’s intellectual bellwether, Ghannouchi’s ideas both reflect and define the ideology of Ennahda.

2011 REVOLUTION

The 2011 revolution serves as a critical pivot point in Tunisia’s political development. As would be expected, the unseating of Zine El Abidine Ben Ali was widely covered by journalists, political scientists, and human rights activists, among others. As a result, there are many resources on the subject. The perspective of journalists provides a view of the revolution on the ground, as it happened. Their perspective offers some broader analysis but specifically helps chronicle the timing of events, the settings of events, etc. The role of food security, youth, the internet, dignity, and ‘the people’ are all discussed in scholarship. Alfred Stepan writes about the Tunisian revolution and the Arab Spring more broadly in the context of democratization theory. Some scholars argued that Islam and Islamists were not noticeable actors in the revolution, and this may have benefitted the revolution. The 2011 revolution captivated the world and set off a series of profoundly significant events throughout the Arab world; it is still widely analysed and discussed.

RECENT TUNISIAN POLITICS

The last area of research is focused on recent Tunisian politics. One core topic is Ennahda in the wake of the revolution, discussed by Kirkpatrick and Lynch, who both highlight Ennahda’s moderate character. Fuller focuses more on the role of Islam in the new political system, and Marzouki focuses on the way Ennahda presented itself to sceptics just after the revolution. Cavatorte and Merone, Marks, and Sharqieh have also investigated Ennahda’s process of reformulation and negotiation during the transition period, with Cavatorte and Merone focusing particularly on Ennahda’s striving for political power. A number of other sources, including periodicals and think tank reports, deal with specific political events and conflicts, such as elections and specific policy debates.

A NEW LITERATURE

Existing literature relevant to this thesis looks either at historical events and processes or current realities. Literature examining both explores history first to form a basis for understanding Ennahda’s position in contemporary Tunisian politics. This thesis focuses on history in order to inform an understanding of Ennahda’s policies and actions while exploring motivations that
demonstrate a continuity between behaviors. Ennahda has undergone tremendous change, as the literature bears out, while also dealing with the same obstacles that it has encountered for decades. By formulating an analytical framework, this thesis reimagines the way literature on this topic may be structured in order to offer a systematic means to understand the Tunisian party’s progress.

This thesis includes a discussion of methodology and then three chapters on the goals-obstacles-steps framework for understanding Ennahda’s behavior. Chapter one details Ennahda’s goals as they currently stand. Chapter two highlights the obstacles that inhibit Ennahda from achieving its goals. Chapter three discusses the steps that Ennahda is taking in order to overcome its obstacles and achieve its goals.

METHODOLOGY
RESEARCH

The goals-obstacles-steps framework is shaped by information and analysis accessed through diverse sources. Chapter one and two include a mix of primary and secondary sources in Arabic, English, and French which largely develop the foundation for the framework by setting the background and context for Ennahda’s behavior. Chapter three also includes primary and secondary sources but is more focused on examining recent behaviors and evaluating their role in the goals-obstacles-steps framework.

Secondary sources were employed in an effort to identify what is driving Ennahda’s goals, obstacles, and steps; such sources include books and journal articles about a wide range of topics including Tunisian history, Ennahda’s history, Rached Ghannouchi, and more. Special emphasis was placed on identifying primary sources such as interviews, speeches, party statements, government documents, election data, books and articles written by relevant figures, and other miscellaneous items. Classified U.S. Department of State documents accessed via WikiLeaks provided insights into the American understanding of Tunisian politics. Statistical information and quantitative data also supported qualitative trends. In order to address bias in sources, authors’ affiliations are mentioned where applicable, and statements of opinion from sources are limited to the extent possible.

There are several limitations in this research process. One constraint is that Tunisia’s political situation is still in transition. Research was conducted throughout 2018 and into 2019, and the framework developed in this paper was tested with behavior between 2011 and 2018. With any luck, this framework will prove robust and generally useful for understanding Ennahda’s actions after 2018, but this is impossible to predict. Another limitation is that field research was conducted only informally and while informing the development of this thesis, does not play an explicit role in this paper.

FRAMEWORK

The goals-obstacles-steps framework examines behavior and actions of the Ennahda party holistically. The framework was developed by the author of this thesis through consideration of major party priorities as well as the factors tied to Ennahda’s political history, that affect the achievement of those priorities. It was initially inspired by the scholarly exchange between Monica Marks and Sayidda Ounissi discussing Ennahda’s strategic responses to isolated events.43 By using historical context for the purposes of evaluating contemporary action, the framework more easily allows one to draw direct connections between modern-day political strategy and the context that influences that strategy. The goals-obstacles-steps framework may be adapted to evaluate the actions of other political decision-makers, so long as historical context is explicitly accounted for. The iteration of the framework discussed in this paper is Ennahda’s goals, obstacles, and steps as listed on the following page. Those specific goals, obstacles, and steps do not apply to Ennahda over the course of its entire history because Ennahda and Tunisian politics have transformed significantly over time. This iteration of the framework, then, is suitable for analyzing the party from the end of 2014, specifically after the 2014 legislative elections, to the end of 2018. While it is possible to look at motivating factors for each action and generate explanations for policy choices or party tactics, this thesis presents a framework that can be used in the present day to contextualize. This is an analytical model, not a predictive model, in that it serves to set existing actions within the broader context of Ennahda’s political efforts. Essentially, the framework requires an action or decision by Ennahda. This action can be then be viewed analytically as falling into one or more of Ennahda’s steps toward one or more of its goals, often in the interest of overcoming one or more of the obstacles to those goals.

For example, a possible use for this framework is to examine the reasons for which Ennahda endorsed
a Jewish candidate in the Monastir governorate in the 2018 municipal elections (discussed in Chapter 3). A possible assessment would be that this was part of Ennahda’s efforts towards moderation and to move away from political Islam (steps) and overcome the effects of vilification (obstacles) in order to ultimately win elections (goals). This is an example and not a definitive explanation, but it demonstrates a formulaic method to understanding the essential rationale for Ennahda’s decision-making.

Figure 1

CHAPTER 1: ENNAHDA’S GOALS

Tunisia today faces a number of complex issues that threaten the future of its fledgling democracy, and each party has its own priorities when confronting these issues. Ennahda’s explicit and implicit goals are rooted in basic democratic strategies and its unique history, making its vision distinct from those of other parties. Ennahda’s goals, in no particular order, are to ensure a stable Tunisian democracy, win political influence through elections, and attract western support for Tunisia. These goals were determined independently through the consideration of existing literature. In light of them, this thesis will go on to profile the obstacles that inhibit Ennahda from achieving these goals and then detail the strategies that Ennahda is adopting to overcome those obstacles.

A. MAINTAINING STABILITY

Before the 2011 ousting of Ben Ali, Tunisians experienced fifty-five years of fairly predictable politics. While there were occasional changes—such as the few odd elections which included opposition parties—the political system generally remained the same. However, since January 2011, Tunisian politics have lacked consistency. Politicians continue to define the details of Tunisian democracy through slow deliberation and debate. Meanwhile, the Tunisian people face economic difficulties that distract them from politics. Ennahda desires stability because Tunisians, accustomed to predictable politics, have little patience for a stalling democracy which fails to improve their economic situation. For Ennahda, whose members were frequently arrested, imprisoned, and exiled under Bourguiba and Ben Ali, maintaining a stable democracy is a matter of existential concern. When asked about Ennahda’s top political priority in 2011, Ghannouchi told Foreign Policy that it was to “guarantee that dictatorship will not return to Tunisia.”

Democracy provides for diverse political voices, preventing one party or ideology from seizing sufficient power long enough to repress its opposition. In other words, democracy ensures that Ennahda will have the right to a political voice. Overall discontent with democracy, then, would concern Ennahda. Ennahda’s goal of stability is not unique. Many parties seek to solidify Tunisian democracy by demonstrating to Tunisians that democracy can solve their problems. The main problems on the table have to do with the Tunisian economy. In 2017, youth unemployment reached 36.3 percent. This grew from the 2010 figure of 29.5 percent. In 2011 the rate spiked dramatically to 42.5 percent. In 2018, Tunisian inflation reached 7.8 percent, its highest point in thirty years. Labor unions, including the UGTT (Tunisian General Union of Labor), have used massive strikes to protest low wages. During this economically tumultuous time, the International Monetary Fund (IMF) urged Tunisia to implement austerity measures in exchange for loans. Such cuts have led to further public backlash and a growing number of public protests. Ghannouchi argues that economic problems should not be overstated, telling Noureddine Jebnoun in 2014, “As for the economy, the country’s economy is not as bad as it is being portrayed in the media. Because of the freedom of expression, everyone is talking about the country’s economy. The bad economic situation of Tunisia under the previous regime in its last years was not better than the current situation.” It does not follow that Ennahda is content with the state of Tunisia’s economy, but Ghannouchi argues that public outcry over economic concerns is counterproductive and that Tunisia’s economic problems are linked more to European economic struggles (due to cross-Mediterranean economic linkages) than domestic shortcomings. Urging calm, Ghannouchi states “one should take into consideration that this is a transition period and the purpose of it is to lay the foundations for a healthy democracy.”

Ennahda’s vision for a healthy Tunisian economy is vague. The party’s governing document, the
Basic Law, calls for “a strong and integrated national economy” and broader integration with North Africa, the Middle East, the Islamic world, and the global economy. In 2011, Ennahda proposed policies to confront unemployment, the post-revolution drop in tourism, corruption, and other concerns. According to Giulia Cimini, the ambition of Ennahda’s economic plans is hampered by the need to play politics and ensure political stability:

“great importance […] is constantly given to fairness, the fight against corruption and social justice in the overall platform as well as in the public discourse of party’s members. In theory, these are the core principles around which Ennahda builds its strategy and discourse. In practice, the necessity to make or consolidate alliances to secure political stability brought the party to take on more nuanced policies, inspired by pragmatism and compromises, policies that place themselves more in continuity with the past instead of being innovative or structurally new.”

Political stability in which alliances are more solidly founded may then be necessary in order for Ennahda to introduce adequate economic reform. Yet, for Ennahda to muster the political influence required to be part of a stable system, it must first successfully hold democratic elections in Tunisia.

The goal of national stability is, as previously stated, not unique to Ennahda. Its allies and opposition alike all seek stability. What is important is not that this goal is unique, but rather that it is central to understand Ennahda’s behaviour. Simply seeking stability in theory is insufficient, as there are real problems in Tunisia’s current political and economic climate that political parties are trying in different ways to address.

B. WINNING ELECTIONS

The Ennahda party is a core force in Tunisia’s newly established democratic system. It has a long history and boasts a long-standing, countrywide network of supporters; these advantages allow the party to be quite successful in contemporary Tunisian politics.

On October 23, 2011, Tunisia held an election for its National Constituent Assembly. Ennahda dominated the election, winning 89 of 217 seats and more votes than the next eight parties combined. Ennahda also “won at least two seats in every Tunisian district and secured nearly five times the popular vote of the second-place CPR [Congress for the Republic].” This chart shows the results of the 2011 election. Ennahda’s lack of is notable; the post-revolution fervor led to the creation of many political parties. The party’s most significant advantages were its existing organizational structure and established countrywide support base, and Ennahda’s longtime vilification by the Ben Ali regime made it an obvious pick for voters looking for an anti-regime party. Still, Ennahda opted to sacrifice its dominating victory and establish a coalition government with Ettakatol and CPR. Commonly known as the Troika government, the coalition was aimed to validate the democratic system. Ghannouchi stated that Ennahda would have formed a coalition even if it received a majority of votes (it merely received a plurality).

Between the 2011 and 2014 elections, several major events impacted the Tunisian political scene, most notably the passing of the Tunisian Constitution on February 10, 2014. Another notable event, which established a political status quo that largely persisted, was the founding of the Nidaa Tounes (meaning Call of Tunisia) party in 2012 by Beji Caid Essebsi (b. 1926). Nidaa was organized as a so-called big tent secularist party, an amalgamation of the several small secular parties that emerged in 2011, to defeat or at least prevent political domination by Ennahda. Secularism in Tunisia is a major political force. The legacy of French colonial rule left a culture in some parts of Tunisia, especially wealthy urban areas, that shares some secular characteristics with Europe. Though almost all Tunisians describe themselves as Muslim,
many consider Islam to be a private matter that has no place in governance. Many of these Tunisians also believe that Islam in Tunisian politics would erode Tunisia’s identity and repress its society. These Tunisians identify as secularists, and Nidaa anchored its base on the idea that Islam is part of Tunisia, but it should not be part of Tunisian politics. As a result, the party starkly opposes Ennahda. Composed of officials from the Ben Ali and Bourguiba regimes, leftists, and mainstream secularists, the new party found immediate success in the 2014 legislative and presidential elections. It appealed to voters who opposed Ennahda and its governance between 2011 and 2014. The results of the 2014 elections illustrate the success of Nidaa Tounes’s campaign. The party won a plurality of votes in the legislative elections, and its leader Essebsi, won the presidency.61

These Tunisians identify as secularists, and Nidaa anchored its base on the idea that Islam is part of Tunisia, but it should not be part of Tunisian politics. As a result, the party starkly opposes Ennahda. Composed of officials from the Ben Ali and Bourguiba regimes, leftists, and mainstream secularists, the new party found immediate success in the 2014 legislative and presidential elections. It appealed to voters who opposed Ennahda and its governance between 2011 and 2014. The results of the 2014 elections illustrate the success of Nidaa Tounes’s campaign. The party won a plurality of votes in the legislative elections, and its leader Essebsi, won the presidency.61

What do these results reveal about Ennahda’s goal of winning elections? In 2011, in the wake of revolutionary fervor, Ennahda benefitted from its widely-known opposition to Ben Ali and a weak field of political opponents. By 2014, an opposition party emerged with a charismatic leader who modeled his image on that of Bourguiba (who secularists had come to see as a symbol of Tunisian progressivism and secularism). This rendered Ennahda’s opposition to Ben Ali less effective and shrunk the gap in party resources between Ennahda and its opposition. With the rise of Nidaa came a new, largely two-party system in Tunisia. This was confirmed by the 2018 municipal elections in which Ennahda and Nidaa both significantly outperformed all other political parties.62

C. ATTRACTING WESTERN SUPPORT

Tunisia, unlike some of its neighbors, is not wealthy in resources. It is also much smaller and less politically influential than some of those same neighbors. As a result, only a few wealthy nations make it a priority to support Tunisia economically or otherwise. This may have benefitted Tunisian revolutionaries in 2011, as they did not compete against a strong military or foreign powers seeking to preserve the status quo. However, it has proved consequential in the years after the revolution. Taking real steps to fix the faltering economy, discussed previously, seems to be political quicksand; support from the International Monetary Fund, for example, comes with significant stipulations which have led to public outrage.

In 2019, Sarah Yerkes from Carnegie Endowment for International Peace asked David Rubinstein, U.S. ambassador to Tunisia, about U.S. economic assistance to Tunisia. Rubinstein gave a vague answer, emphasizing the U.S. “commitment” to economic assistance and encouraging Tunisia to curb corruption and political “competitiveness.” U.S. records show that U.S. commitments and appropriations are not leading to adequate assistance. In 2013, for example, the U.S. appropriated $47.2 million to Tunisia, but only ended up spending $15.4 million (32.6 percent). In 2016, the U.S. appropriated $141.85 million and only spent $21.63 million (15.2 percent).66

According to an email leaked by WikiLeaks in 2016, the U.S. Department of State established a “Democracy Transition Team for Tunisia” in the months after the 2011 revolution to work on “education, elections, economic empowerment, media engagement and training, women’s issues, and, launching Peace Corps in Tunisia.” There was also discussion about supporting the Tunisian SME (small and medium enterprise)
sector. However, voices in the State Department were reportedly anxious and skeptical about the efforts, as a transition team was a “new idea for the Department.”

Though efforts appeared to be significant in the run-up to the 2011 Tunisian elections, such efforts did not continue beyond Ennahda’s victory and the formation of the coalition Troika government (comprised of Ennahda, Ettakatol, and CPR).

In 2016, when asked if the West was offering support to Tunisia, Ghannouchi responded bluntly, saying “Not enough. I think the West gave more support to democracy in Eastern Europe than Tunisia …” Specific examples of this kind of support were given by Said Ferjani, a senior Ennahda member. These included increasing trade, financing infrastructure projects, and encouraging consensus-building among political parties. Ennahda’s secular counterparts also seek support from the West, but from different countries than Ennahda. Ennahda seems more prone to appeal to the U.S. for assistance than European countries, possibly because, as a 2013 joint report from Hivos, Arab Forum for Alternatives, and FRIDE states, “the most substantial political criticism from Tunisian actors consisted in the EU and most of its member states having adopted an approach that was too biased towards secular political forces, to the detriment of Islamists.”

Ennahda seeks Western support in part because seeking support from Arab Gulf countries is politically volatile due to regional tensions. For example, Nidaa Tounes gained support in 2014 from Saudi Arabia and the United Arab Emirates, while their rivals Qatar and Turkey supported Ennahda. When Nidaa and Ennahda formed a coalition government after the 2014 elections, Saudi Arabia and the UAE were perturbed, according to Asma Ajroud. The UAE even offered Tunisia $5 billion to $10 billion if Nidaa would break off the coalition, according to Middle East Eye. In November 2018, Mohamed Bin Salman of Saudi Arabia visited Tunisia for a widely-protested meeting with President Beji Caid Essebsi, a member of Nidaa Tounes; Bin Salman extended Tunisia a loan of $500 million less than a month later. Ennahda seems hesitant to engage outright with the Arab world, as it is an ever-contentious issue. The party also avoids associating with Qatar and Turkey because those countries have demonstrated support for the Muslim Brotherhood, and Ennahda is wary of being associated with the Muslim Brothers (see Chapter 2b and 3b). Ennahda’s appeals to the West for support rather than reaching out to the Arab world (particularly the Gulf) also make sense given Ennahda’s commitment to Tunisian liberal democracy. Western countries are more likely to couple financial support with political support than non-democratic Gulf states.

SUMMARY
These objectives of achieving stability, winning elections, and attracting Western support are fundamental goals that Ennahda holds as crucial to establishing a strong post-revolution Tunisia. It is difficult to argue that Ennahda has achieved any of its objectives. Clearly, public trust in the Tunisian government is lacking, and economic problems put the stability of the country’s democracy at risk. Despite holding a strong core voter base, Ennahda also has a great deal of critics, competes with sizeable rivals, and faces an increasingly apathetic voter pool when it comes to elections. Ennahda’s leadership is also clearly dissatisfied with the level of assistance coming from Western countries, though, to the author’s knowledge, it has not requested specific amounts of support.

CHAPTER 2: ROOTS OF OPPOSITION
Ennahda faces unique hurdles to achieving its goals. These obstacles are rooted in Ennahda’s history, its perceived ideological identity, and its behavior. Ennahda’s chief obstacle to achieving its three main goals (national stability, election victories, and Western support) is widespread apprehension among secular Tunisians. This apprehension stems from three main sources: Ennahda’s repression under Bourguiba and Ben Ali, association with political Islam, and perceived corruption of the goals of the 2011 revolution.

A. VILIFICATION BY BOURGUIBA AND BEN ALI
The Ennahda Party can trace its roots back into Tunisia’s history. Even from 1971 to 1987, when it existed only as MTI [Mouvement de la Tendance Islamique, or Islamic Tendency Movement], the organization grew rapidly and gained support from conservative Tunisians looking to increase the role of Islam in Tunisian civil society. Owing to France’s colonial imprint and secularism, as well as Bourguiba’s policies of Tunisian secularism, the country lacked a religious culture that marked nearby countries like Egypt and Syria. It was in those countries where Rached Ghannouchi learned about Arab nationalism and then Islamism through his travels in the 1960s.

Even at the movement’s nascent stage, it ran
into some trouble with Bourguiba’s government. But according to leftists, who also opposed Bourguiba, the government was tolerant of the Islamists. Hamdi puts forth that “it may be safe to assume that the authorities did not initially realise the potential threat of these young Islamist activists.” It is also likely that Ghannouchi and other young Islamists did not express any political goals. The movement grew secretly. When its true size and ambition was discovered in 1980, months after the Iranian Revolution, Ghannouchi declared the MTI’s desire to form a political party. Subsequently, the MTI’s relationship with the Bourguiba government turned oppositional, and Bourguiba began to take punitive measures to repress the MTI’s expansion. Though repression benefitted the movement in some ways, including earning it some sympathy, it also began a pattern of government vilification that continues to impact public opinion of Ennahda.

According to Mohamed Elihachmi Hamdi’s book *The Politicisation of Islam*, in the early 1980s “the government was adamant that the MTI was a blatant threat to society,” and government media “denounced ‘the entire Islamic movement.’” At some point in the mid- to late-1980s, a sub-group known as the Security Group formed within the MTI with the sole purpose of deposing Bourguiba. Plans were complicated when in August 1987 a separate group claiming allegiance to the MTI conducted several bombings in Sousse and Monastir. Though Ennahda denounced the attacks and denied knowing the group or its leader, “The entire state apparatus was mobilised in the battle against the MTI on direct orders from [Bourguiba].” The Security Group’s plan to remove Bourguiba remained in place and was set to occur in early November 1987. When the plan leaked, Ben Ali, then the Minister of the Interior, stepped in himself to unseat Bourguiba before any action was taken by the Security Group.

Ben Ali’s accession to the presidency was an optimistic time for the MTI; the new president freed hundreds of political prisoners in 1988 and early 1989 and, according to sociologist Elbaki Hermassi, quoted by Rory McCarthy, “aligned himself with Tunisia’s ‘Arab-Islamic identity.’” He opened the political process to new parties in 1988 and the MTI applied as Ennahda, or the Renaissance, as it was not permitted to use the word Islam in its name. It declared its goal to foster a “civil democratic state, rather than an Islamic state.” Throughout the campaign for the 1989 election, in which Ennahda’s members ran as independents since their party application had been rejected, everything seemed above board until the election results were announced. Democratic Constitutional Rally (RCD), Ben Ali’s party, was awarded all 141 seats in the parliament, despite results that showed a level of popular support for Ennahda which “dazzled its leaders.” This kicked off a period of hostility between Ennahda and Ben Ali. Ennahda’s impatience after glimpsing the daylight of political involvement led to “strident public statements,” a number of protests, and even “a rare act of violence,” Ben Ali responded by arresting and imprisoning thousands of Ennahda members and by outlawing the party. Though some members, including Ghannouchi, avoided imprisonment by going into exile, the party was crippled. Without any real opposition, Ben Ali consolidated his power, tortured Ennahda prisoners, and continued to paint the organization as an imminent threat to Tunisia’s stability and wellbeing. In 1991, Ben Ali’s government claimed that it discovered an extensive Islamist coup plot and arrested more than 300 people, but it was later discovered that the plot was a fabrication. With a suffocating police presence came social ostracization as Ben Ali’s vilification efforts took hold. From secular society, which did not experience the exclusion faced by Ennahda and which feared Ennahda would “alter the secular nature of the state,” came a “tacit consensus for repression.” As a result, Ennahda became increasingly sidelined and marginalized. Though the movement endured, in large part thanks to the wives and families of the imprisoned party members, significant police persecution made life difficult for Ennahda members.

Ben Ali also used Tunisian media to control information about Ennahda, and even took steps to influence perceptions of Ennahda and his own regime abroad. According to a classified 2005 memo from the U.S. Embassy in Tunisia, “Ben Ali has used the Ministry of Interior to ensure religious elements never again gained such popular support.” When Ghannouchi was asked in late 2011 about people who feared Ennahda, he said “The people who do not trust us, it’s normal, because for twenty-two years they have been subject to propaganda from Ben Ali which has discredited us, and made people fearful of us.”

B. ASSOCIATION WITH POLITICAL ISLAM

In 2005, the U.S. Embassy in Tunisia sent a memo to the U.S. Secretary of State and several U.S. embassies stating, “Many secular Tunisians, especially those of the Bourguiba generation, also completely
reject the possibility of including conservative religious elements in government, as they believe these groups will reverse the social and cultural progress Tunisia has made since independence.100 For secular Tunisians who seek to protect the secular nature of the state from the influence of Islam, political Islam is a real threat. Political Islam, or Islamism, is defined by The Oxford Encyclopedia of the Islamic World as “the views of those Muslims who claim that Islam, or more specifically, the Islamic shari‘ah, provides guidance for all areas of human life, individual and social, and who therefore call for an ‘Islamic State’ or an ‘Islamic Order.’”101 Under Ben Ali, the word Islamist was politicized by the government to “tarnish the image” of Ennahda and other religious groups it felt threatened state stability.102 For this reason, along with broader politicization of the term Islamism globally, the term political Islam became more popular as a distinct term.

Tunisian secularists have long alleged Ennahda’s association with two main actors in the world of political Islam: the Egyptian Muslim Brotherhood and Iran. Though the movement was certainly influenced by foreign Islamists, mostly in its early years, opponents over time have claimed that foreign groups control the Ennahda members or that the party otherwise prioritizes the broader Arab world over Tunisia. These claims persist today. Despite evidence to the contrary, “the dominant lens for analyzing Islamist parties remains subtly shaped by assumptions that Islamists move in lock-step, that they behave cultishly and act unilaterally, and that they pose an inherently greater threat to pluralism than secular, leftist, or otherwise self-avowedly liberal parties.”103 Ennahda is impeded in achieving its goals by its perceived association with the Muslim Brotherhood, Iran, Salafis, and ISIS.

The MTI emerged amidst a resurgent interest in political Islam, and Islam more generally, following the 1967 Arab-Israeli War. Ghannouchi, one of three co-founders of the Islamic Tendency Movement, turned to Islamism from Arab nationalism. Ghannouchi did this while in Syria after spending time with the Syrian Muslim Brotherhood (MB).104 Tunisian Islamists more broadly were influenced by Egyptian Brotherhood books, as well as Algerian Islamist Malik Bennabi.105 In the early 1970s, the Muslim Brotherhood at this time was highly influential across the Arab world, as it was the oldest group advocating for the increased role of Islam in society and politics.106 However, the MTI gradually grew apart from the MB in order to focus on Tunisia’s particular reality and the benefits of democracy.

Despite a fissure between the MB and Ennahda that has widened over time,107 some still claim ties exist between the two groups, with some saying that Ghannouchi is part of the “Global Muslim Brotherhood” or that “Ennahda is the Muslim Brotherhood in Tunisia.”108 When asked about his affiliation with the International Union of Muslim Scholars, an international group headed by “one of the top Muslim Brotherhood ideologues,” Ghannouchi alleged to the Washington Post that the group is not political, possibly meaning that it has no specific political goals.109 The group espouses highly political rhetoric and supported the Arab Spring revolutions.110

Though the Ennahda’s history extends back to the 1960s, it was the Iranian revolution of 1979 that sparked a push for political activism, just as it inspired Muslims around the Arab world to consider the role of Islam in politics.111 It does not appear that Ennahda had any effectual contact with Iran. Yet fear of Iranian influence led Bourguiba, for example, to claim that the 1987 Sousse and Monastir bombings were backed by Iran and spearheaded by the MTI.112

On the Tunisian political spectrum, Ennahda is understood to be conservative, but not as conservative as political Salafists.113 Salafists believe in a short-term strategy of Islamizing politics, often through physical jihad, or struggle. Ennahda has long denounced these groups, especially Ansar al-Sharia, but at different times opponents have claimed that Ennahda is either affiliated with or sympathetic to these groups. Ennahda’s efforts to condemn Salafism following terrorist attacks and assassinations in 2012 and 2013 were criticized by much of Tunisia’s public, along with “many scholars and analysts,” as too weak and slow; these critiques had real consequences for Ennahda in the 2014 elections.114

Although few contend that Ennahda is directly involved in terrorist activities, the acute fear many Tunisians have about Salafist terrorism and radicalization often translates to criticism of Ennahda, since it is seen to be the largest political force representing political Islam in Tunisia. Terrorism in Tunisia raises two primary concerns. The first is concern about security and safety. The second is economic. Terrorist attacks are particularly devastating to the country’s stability given their effect on Tunisia’s fragile tourist industry. Tourism is a crucial sector in the Tunisian economy, but since the revolution tourist numbers have been volatile, due especially to the difficulty of reassuring cruise
Tourism numbers plummeted in the wake of the March 2015 terrorist attack and hostage crisis at the Bardo Museum in Tunis; of the 23 people who died, 19 were international tourists. According to the World Bank, international tourism receipts dropped from $3.042 billion in 2014 to $1.869 billion in 2015 and then to $1.706 billion in 2016.

Perhaps the most sensitive issue related to political Islam in Tunisia is ISIS (Islamic State in Iraq and ash-Sham (or Syria)), an organization aiming to establish a caliphate, or Islamic state headed by a caliph, first in Syria and Iraq, and then globally. International forces have diminished ISIS territory substantially since the group declared a caliphate in 2014, but since that time ISIS has become an agitating topic in Tunisia because of the group’s recruitment efforts. While Ennahda is not affiliated with ISIS, understanding the public, especially secular, consternation over the relationship is absolutely necessary to understand opposition to Ennahda as a symbol of political Islam in Tunisia.

There are more foreign fighters from Tunisia in ISIS than from any other country by a wide margin. According to a report from the National Bureau of Economic Research, there were 6,000 official (7,000 unofficial) foreign fighters from Tunisia in ISIS, the highest amount by over 3,000. As of April 2016, only 970 had returned to Tunisia. Tunisia also sent the highest number of fighters per capita, at 545.5 fighters for every one million total Tunisians. These numbers are shocking to secular Tunisians and cause them to confront a dichotomy in Tunisian society between themselves and religious Tunisians. The issue has even permeated popular culture and become part of the societal narrative, as shown by the solemn 2018 movie called Dear Son, directed by Mohamed Ben Attia, which focuses on the parents of a foreign fighter.

Of those 6,000 Tunisians who joined ISIS, many were women. Tunisian women, according to an investigation by Aaron Zelin, play key roles in ISIS. Back in Tunisia, those used to a more secular society have a difficult time understanding how women born and raised in Tunisia found ISIS appealing, given its hyper-religious message. The notion of *jihad al-nikah* is perhaps the most polarizing issue about Tunisia and ISIS; the equivalent term in English is sexual jihad, and it means the practice of women marrying fighters temporarily to have sex with them to improve morale. The topic is extremely controversial. In 2013, Ottman Battikh, Grand Mufti of Tunisia, stated that the phenomenon was occurring and was promptly dismissed from his position (he was reinstated in 2016). When Minister of the Interior Lotfi bin Jido accused Tunisian women of undertaking *jihad al-nikah* in 2013, it was criticized as unfounded and an attempt to undermine the Ennahda-led government. While the issue became more well-known as more Tunisians joined ISIS, Aaron Zelin states that the government continued to indicate that “no credible evidence has emerged to suggest women joined IS specifically to ‘comfort’ fighters.”

Ennahda certainly has an association with political Islam but interpreting the extent and role of association is highly politicized and personal for Tunisians. Skeptics are also fueled by Western voices, like Martin Kramer of the Washington Institute for Near East Policy. Kramer said of Ennahda’s calls for democracy, “Islamists become the more moderate and tolerant of pluralism the further away from power they are.”

In the wake of the revolution, “many Tunisians predicted that Ennahda, an assumedly transnationally linked Islamist actor, would actively aid – or, at the very least, tacitly abet – the importation and popularization of jihadism and Egyptian-style Islamism in Tunisia.” While evidence for this happening is scant, the fear of Ennahda’s association with political Islam defined many Tunisians’ views of the party, and impelled Ennahda to distance itself from Islam, as will be discussed in Chapter three. Though some opponents of Ennahda believe that its greatest faults lie in this affiliation with the world of political Islam, there are also those who point to domestic issues as sources of frustration with the Renaissance Party.

### C. CO-OPTING THE GOALS OF THE 2011 REVOLUTION

The 2011 Revolution in Tunisia was a pivotal moment in the country’s history. Sparked by Mohammed Bouazizi’s self-immolation in the capital of the interior Sidi Bouzid governorate on December 17, 2010, mass protests across Tunisia against government apathy gradually turned into calls for the deposition of Ben Ali. When Ben Ali fled to Saudi Arabia on January 14, 2011, most Tunisians were surprised, “Given the historical ineffectiveness of Arab publics to effect real change in their governments [...]”. As a result of Ben Ali’s exclusion of Ennahda and Islamists from
In Tunisia's current democratic system, the political opposition to Ennahda that has brewed for decades acts manifest in the ballot box and the streets. Ennahda works to influence the political scene in spite of this opposition, but its ability to achieve its objectives is constrained by the political landscape.

Ennahda's roots stretch back into independent Tunisia's history. To Ennahda's benefit, its long history has solidified its support base over the years. To its detriment, though, Ennahda's long history with police measures, exclusion from elections, and political polarization has also solidified opposition by segments of society. The intransigence of Ennahda's supporters, combined with the popular revulsion against the old regime, has led to the party's exclusion from the electoral process. As a consequence, Ennahda's influence over the political landscape is limited.

Ennahda's annual statements commemorating the revolution's anniversary often reflect on the role of political parties and the need for a united front. The party's reliance on the Muslim Brotherhood in Egypt's 2011 revolution is sometimes cited as a metaphor for the need for a united front and a shared agenda. The January 2018 statement refers to the revolution's role with regards to protecting the revolution's goals by Ennahda, but it is unclear either how the goals of the revolution could be taken as a reclamation (or hijacking) of the revolution's goals by Ennahda, or not stating its commitment to those objectives.

Paradoxically, then, Ennahda must choose between stating its commitment to fulfilling the objectives of the revolution and being accused of co-opting or hijacking them, or not stating its commitment to those objectives and being accused of only pursuing self-serving goals. The party's ability to achieve its objectives is constrained by the political landscape, and its influence over the political scene is limited.

In analyzing party statements, there is no clear evidence that the party or its leader ever truly claim any ownership of the goals of the revolution. Chamouni stated, the revolution cannot be contained or exploited by any political party. Therefore, when asked to fulfill the objectives of the revolution, Chamouni believes political parties have no right to exploit popular revolutions. However, Ennahda's annual statements commemorating the revolution's anniversary suggest that Chamouni believes political parties have no right to exploit popular revolutions. However, Ennahda's annual statements commemorating the revolution's anniversary are more inconsistent on the party's role with regards to protecting the revolution's goals by Ennahda, but it is unclear either how the goals of the revolution could be taken as a reclamation (or hijacking) of the revolution's goals by Ennahda, or not stating its commitment to those objectives.

In Tunisia's current democratic system, the political opposition to Ennahda that has brewed for decades acts manifest in the ballot box and the streets. Ennahda works to influence the political scene in spite of this opposition, but its ability to achieve its objectives is constrained by the political landscape. Ennahda's reliance on the Muslim Brotherhood in Egypt's 2011 revolution is sometimes cited as a metaphor for the need for a united front and a shared agenda. The January 2018 statement refers to the revolution's role with regards to protecting the revolution's goals by Ennahda, but it is unclear either how the goals of the revolution could be taken as a reclamation (or hijacking) of the revolution's goals by Ennahda, or not stating its commitment to those objectives.

Paradoxically, then, Ennahda must choose between stating its commitment to fulfilling the objectives of the revolution and being accused of co-opting or hijacking them, or not stating its commitment to those objectives and being accused of only pursuing self-serving goals. The party's ability to achieve its objectives is constrained by the political landscape, and its influence over the political scene is limited.

In analyzing party statements, there is no clear evidence that the party or its leader ever truly claim any ownership of the goals of the revolution. Chamouni stated, the revolution cannot be contained or exploited by any political party. Therefore, when asked to fulfill the objectives of the revolution, Chamouni believes political parties have no right to exploit popular revolutions. However, Ennahda's annual statements commemorating the revolution's anniversary suggest that Chamouni believes political parties have no right to exploit popular revolutions. However, Ennahda's annual statements commemorating the revolution's anniversary are more inconsistent on the party's role with regards to protecting the revolution's goals by Ennahda, but it is unclear either how the goals of the revolution could be taken as a reclamation (or hijacking) of the revolution's goals by Ennahda, or not stating its commitment to those objectives.
goals is limited by its opponents’ efforts. Therefore, Ennahda is taking steps to allay opposition and present an image of itself as a party that accords with secular visions of Tunisia as well as more religious visions of the country.

CHAPTER 3: OVERCOMING OBSTACLES
Ennahda has undergone profound changes in the past several years. As Tunisia’s political system settled into a two-party structure during and after the 2014 elections, Ennahda’s focus narrowed to competition with one sizeable opponent: Nidaa Tounes. After the 2014 elections the Ennahda party entered the current period of analysis. Before this point, the iteration of the goals-obstacles-steps framework discussed in this thesis was not applicable. The goals of the party were the same, but the contextual dynamics of those goals changed after the 2014 elections. The dynamics of stability changed with the ratification of Tunisia’s constitution and a turnover in power. The dynamics of elections changed with the emergence of a two-party system. Finally, the dynamics of Western support changed with a secular party in power. While the obstacles of vilification and co-opting of the 2011 revolution stayed largely the same, the world of political Islam continued to change, most notably with the rise of ISIS.

In this new context there are three principal steps that Ennahda is taking in pursuit of its goals. These steps are emphasizing the party’s Tunisian roots, moving away from political Islam, and adopting moderate governance strategies. Note that some examples used in this chapter to support the steps occurred prior to the 2014 elections because, as noted in the previous paragraph, the current goals and obstacles are not altogether new.

A. EMPHASIZING ENNAHDA’S TUNISIAN IDENTITY
According to Sayidda Ounissi, an Ennahda legislator, accusations that Ennahda is the Tunisian branch of the Muslim Brotherhood are false and have led to “misinterpretation.” While the Brotherhood and its publications were “certainly important at the intellectual level” to the MTI in its early days, Ounissi states that the founders of MTI/Ennahda had a “very nationally grounded trajectory.” Still, many do not believe that Ennahda’s ideology and allegiance are Tunisian. Ennahda is working to emphasize its Tunisian roots and identity whenever possible by re-evaluating its stance toward Bourguiba and the Personal Status Code and by highlighting its opposition to terrorism.

BOURGUIBA
The memory of Habib Bourguiba, the first president of independent Tunisia, is prominent in Tunisia today. Ben Ali, after deposing Bourguiba, consolidated his new authority by removing evidence of his predecessor from public spaces. He changed street names, public square names, and even removed a prominent statue of Bourguiba from the main street in downtown Tunis. The removal of Ben Ali in 2011 allowed for renewed expressions of nostalgia for the Bourguiba era. As a 2016 report puts it,

“For the majority of commentators and historians, and within Tunisian national memory itself, Bourguiba remains a key figure in the recent history of Tunisia, and seems to have marked the entire period from 1927 to 1987 and even beyond. To many, he is still the main character of the national struggle against French colonisation, the leader of great social reforms in Tunisia, the liberator of women. His legacy runs deep.”

Many older secular Tunisians remember Bourguiba fondly, while religious elements in the country see such nostalgia as overly romantic. Bourguiba was an avowed secularist, famously drinking orange juice on television during the fast of Ramadan in 1964. He cracked down on all opposition groups, including both Islamists and leftists, and handled the Tunisian economy fairly poorly. Although secularists saw him as liberating Tunisia, Ghannouchi argued in 1994 that “independence achieved by Bourguiba was not […] a victory for the local cultural identity against Westernization. It was, in fact, a resumption of […] the destruction of the civil society which was started by the colonizer.”

Bourguiba’s memory is more distant than that of Ben Ali’s, but for Ennahda, and Ghannouchi in particular, it remains intact. Bourguiba imprisoned Ghannouchi multiple times and, prior to the 1987 Ben Ali coup, even tried him for the death sentence. For secular Tunisians, many of whom were not alive for his time in office, Bourguiba’s renewed and romanticized memory was galvanizing. This memory led many to vote for Beji Caid Essebsi, Bourguiba’s former Minister of Foreign Affairs. As a result, Ennahda’s leadership was compelled to weigh its memory of Bourguiba against the public’s memory. Strategically, as part of the broader step to emphasize the party’s
accomplishments, while staying shy of outright praise. Ghannouchi described the party’s stance to Middle East Eye Editor in Chief David Hearst in 2016, saying “You cannot airbrush [Bourguiba] out of our history. He guided a national movement which liberated Tunisia. I cannot negate that. It is a reality.”

Additionally, a 2016 report on the collective memory of Bourguiba stated that “the return of the statue of Bourguiba to downtown Tunis has not raised large protests among Ennahda party members, although it may have shocked its deeper base.” While not the most cordial position, Ennahda’s acceptance of Bourguiba’s impact allows the party to integrate its view of Tunisian history with the secular view. It is in this acknowledgement that Bourguiba is both part of Tunisia’s history and part of Ennahda’s history, the party ties that its history to Tunisia. By blurring history in this way, Ennahda also seeks to diminish the effects of Bourguiba’s information campaign against the MTI.

PERSONAL STATUS CODE

One element of Tunisian politics that many secularists regard as fundamental to Tunisia’s unique stature in the Arab world is the 1957 Personal Status Code. The brainchild of then new president Bourguiba, the Code included landmark rules on women’s rights, such as a polygamy ban, a formal divorce process, and a mutual-consent marriage stipulation. The MTI believed the Personal Status Code was un-Islamic. Besides constraining commonly understood Islamic law, the Code also symbolized Bourguiba’s broader mission of reform. It also signaled open adherence to Western societal values at the expense of traditional values, which conservatives, including Ennahda, resisted.

Given the Personal Status Code’s widely-known emphasis on gender equity, the MTI’s opposition to the Code made the MTI an easy target for secularists. The issue was used as an arrow in the information campaign quiver of the Bourguiba and Ben Ali regimes, as Ghannouchi told the Financial Times in early 2011: “The government used to always say to frighten people away that (the Islamists) will take away the rights of women to frighten people away.”

In response, the movement took clearer steps to oppose the Code’s Western characteristics, instead of its feminist ones. In 1988, Ghannouchi spoke on the deficiencies in the Islamic movement, and specifically about this issue. He bemoaned Islam’s “lack of awareness and insensitivity toward the oppression, degradation, abasement, restrictions of their horizons and roles in the life of society that women have endured during the long centuries of decline.” He went on to say that “It was only natural for women to be influenced by the enticements of the West, because she was already suffering under the yoke of an oppressive, false Islam,” but arguing that “The Muslim woman went from being the slave of the man and the family to being the slave of the capitalist, of political and media institutions, selling her body[…]” Despite MTI/Ennahda’s long-time opposition to the Code, the Code persisted, coming to symbolize the progressive values of Tunisia for the country’s secularists. Now facing accusations of anti-feminism, Ennahda had to re-strategize.

To reassure the secularists, Ennahda took many opportunities to acknowledge its respect for the Personal Status Code as cornerstone of Tunisian society, as part of the broader step of emphasizing the party’s Tunisian identity. Ghannouchi told FT in 2011 that Ennahda’s members “all recognise and accept the personal status code and will not cancel it or reject it.” He also reminded FT that he had said the same in 1988.

As the Ennahda-led coalition controlled the National Constituent Assembly from 2011 to 2013, the Assembly worked to craft a constitution for post-revolutionary Tunisia. The Personal Status Code and women’s rights again reared their heads. In a draft of Article 28 of the Constitution, which addressed gender roles, Ennahda pushed for language that said men and women are “complementary,” while secularists insisted on “equal.” This debate gained much media attention and secularists painted the term “complementary” as misogynist and backwards, though Ghannouchi insisted that the term meant men and women complement each other. Fear among secular Tunisian women that they would be oppressed under an Islamic government made the issue particularly fiery. Ultimately, Ennahda conceded and agreed to the term “equal,” sending the message that the party would accept non-traditional rules if those rules accorded with the Tunisian identity. In a 2015 interview with Anne Wolf, Ghannouchi said of women’s rights that they “are nowadays part of the Tunisian identity. So Tunisian Islamists try to adapt Islam to [this] reality.”

In 2016, according to Monica Marks, “Ennahda’s policy priorities have shifted from defending Arab-Islamic identity to addressing material concerns of average Tunisians, including security, job creation, administrative inefficiency and corruption.” This shift is part of the broader step of emphasizing Ennahda’s Tunisian identity. In another instance of choosing Tu-
nisiar priorities over broader Arab-Islamic priorities, the party agreed, after intense internal and public debate, to omit the word *shari’a* (Islamic law) from the Constitution’s Article 6 because of the term’s ambiguity and controversial nature.157 While this omission led to mass protests and a degree of party fracturing,158 it reinforced the willingness of Ennahda to compromise, in the name of consensus, on issues that related to perceptions of Tunisia’s particular identity. This willingness is rooted in Ennahda’s flexible understanding of Islam which allows Ennahda, according to Ghannouchi, “to rethink Islam in Tunisian manners,”159 and offer “compromises in terms of the constitution so that it will represent all Tunisians.”160 It is worth noting that the willingness to compromise with secular viewpoints is also useful in trying to attract support from moderate secular voters and secular Western countries.

**ANTI-TERRORISM**

Criticism of Ennahda stems in part from concerns about the party’s relationship to terrorists and violent actors. Events during the party’s 2011 to 2014 coalition government included a September 2012 attack on the U.S. embassy in Tunis (response to American film *The Innocence of Muslims*) and assassinations of well-known secular leaders Chokri Belaid and Mohamed Brahmi in February and July 2013, respectively (both attributed to Salafist group Ansar al-Sharia).

The assassinations erupted tensions between Ennahda supporters and critics. After Belaid’s assassination in February, the Ennahda Prime Minister announced that he would dismiss the government, but Ennahda leadership convinced him to reverse his decision. A further series of mass protests, in opposition and in support of Ghannouchi and Ennahda led to a political crisis. After Brahmi’s assassination only six months later, political chaos put the country at a breaking point and a coup seemed probable. According to Robert Worth, Ghannouchi decided to shoulder the blame for Ansar al-Sharia’s rise and the assassinations, siding with Tunisia’s democracy over Islamism, to minimize violence between Islamists and secularists.161 At a rally only nine day after Brahmi’s assassination, Ghannouchi told a crowd of supporters “‘National unity is the highest goal, and we will need to make sacrifices for this unity. […] Tunis is older than Ennahda, Tunis is older than all of us.’”162 In late August, Ennahda declared Ansar al-Sharia a terrorist group, but by then it was too late to regain control. The party had made too many mistakes and been too lenient with Salafists and Islamists since the revolution.163 To the dismay of many party leaders, Ghannouchi signed an order resigning the party in January 2014 in favour of a technocratic government.164

Although this series of events traumatized the party, it also led Ennahda to strengthen its public opposition to terrorism, a stance it began to communicate more broadly after it left power in 2014. The true test of the party’s renewed position on extremism came in 2015, when another series of tragedies struck Tunisia. These included an attack and hostage crisis at the Bardo National Museum in March (attributed to Al-Qaeda in the Islamic Maghreb), an attack at a hotel in Sousse in June (generally attributed to Ansar al-Sharia), and the bombing of a presidential guard van in downtown Tunis in November (attributed to ISIS). Though it did not control the government, Ennahda responded immediately to each of these attacks with calls for stricter security, a far cry from how it might have responded in 2013.165 In early 2016, Ghannouchi distanced Ennahda from ISIS with an article in *The Atlantic* entitled “Fight ISIS With Democracy,” in which he outlined the principles needed to fight ISIS locally, including freedom, economic empowerment, quality education, and more.166

In his *Atlantic* article, Ghannouchi argues that “the fight against ISIS must be not only a fight against something, but a fight for something (emphasis not added).”167 He promotes the notion of Tunisia’s unique position and influence in the discourse on democracy in the Arab world. This argument illustrates that Ennahda chooses to ally with secular Tunisians in the country’s anti-terror movement in an effort to emphasize the party’s Tunisian identity and roots and overcome the obstacle of affiliation with political Islam.

**EMPHASIZING ENNAHDA’S TUNISIAN IDENTITY: SUMMARY**

From 1956 to 2011, Tunisia was a largely inhospitable environment for Islamists. As MTI/Ennahda grew, it looked to external models and avoided associating itself with the Tunisian government. In post-revolution Tunisia, this has exposed Ennahda to accusations of foreign allegiance, complicity toward terrorism, and merely superficial coherence with the Tunisian identity. As part of the step of emphasizing its Tunisian roots, Ennahda tolerates the public’s renewed memory of Bourguiba, demonstrates a willingness to compromise on issues related to the Personal Status Code, especially women’s rights, and shows a desire to join with secularists to oppose terrorism. This step

153
is important as it serves to counter allegations that Ennahda is not fundamentally Tunisian. The party is also taking radical steps to dispute these allegations by publicly moving away from political Islam and trying to “integrate Islam into Tunisian realities,” or make it align with popular notions of Tunisian identity. In the past, Ennahda acted with the belief that popular notions of Tunisia’s identity could, and should, incorporate political Islam. The party’s recent movement away from political Islam indicates that the party’s leadership no longer believe that popular notions of Tunisian identity will come to incorporate political Islam.

B. MOVING AWAY FROM POLITICAL ISLAM, OR DEPOLITICIZING ISLAM

Following the 2014 legislative and presidential elections, Ennahda began to reevaluate its relationship with Islam. Many of Ennahda’s Bourguiba-era critics opposed MTI’s politicization of Islam, but MTI still made its name as an Islamic, or Islamist, movement; that identity persisted even after the name change to Ennahda. Reevaluating the party’s relationship with Islam, then, meant formally confronting a long-standing and core pillar of the party, but party leadership determined it was necessary in order for Ennahda to achieve its goals. This shift is so fundamental that it can be considered a revision of the party’s identity. It is not the first identity revision the party has undergone, but as the post-revolution system settles this revision may have long-lasting effects. Knowing that Ennahda and its leader Rached Ghannouchi are willing to reevaluate the role of Islam in the party’s vision for Tunisia is crucial.

MUSLIM DEMOCRATS

From the MTI’s creation, there has been controversy around its use of the term “Islam” in political messaging. As almost all Tunisians, after all, are Muslim, the suggestion that one party could represent a distinctly Islamic vision irritated many people. When the MTI applied to be a political party in 1988, it was required to remove the word Islam from its name; it became Ennahda, or the Renaissance party. Still, Islam remained the defining characteristic of the party with the term Islamist used, to this day, to refer to Ennahda. Ghannouchi, as a long-time proponent of democracy, often discussed the bond between Islam and democracy, both theologically and politically. This argument persisted after the revolution. Ennahda, as an Islamic party seeking power in a democratic system, promoted a vision of a pluralistic democracy cohesive with Islamic values. In 2012, Ghannouchi addressed the Centre for Islam and Democracy in Tunis and argued for the compatibility of Islam and democracy, saying “should Muslims be in need of laws, the democratic mechanism is the best embodiment of the Shura (consultation) value in Islam.” Even then, though, Ghannouchi hinted at the party’s ambivalence toward the issue of political Islam, saying “The primary orbit for religion is not the state’s apparatuses, but rather personal/individual convictions.” In the same speech he discussed a notion that he had proffered since at least 1994, a definition of secularism that allows for Ennahda’s participation in government. He said, in 2012, “the greater part of the debate taking place nowadays in our country is a misunderstanding of such concepts as secularism and Islam. We demonstrated that secularism is not an atheist philosophy but merely a set of procedural arrangements designed to safeguard the freedom of belief and thought.”

At its May 2016 congress, the party took a momentous step to formally move away from political Islam by decoupling its political efforts and its religious proselytization efforts. According to Monica Marks, the changes were “tweaks” undertaken in the act of “formalizing long-brewing trends within the party,” but as part of the broader process of influencing perception of Ennahda externally, these changes mark a bold step. Prior to 2016, party leaders could serve as leaders in both Ennahda as well as civil society organizations. Ennahda leaders were also permitted to preach; those allowances were stripped at the party congress. The political party that remained was functionally secular, by the “procedural arrangements” definition Ghannouchi referenced in 2012. This move was not a complete secularization, it is worth noting, as Ghannouchi said “We didn’t leave Islam. Islam is our reference point.” The party did not de-Islamize, but rather began a process of de-politicizing its relationship with Islam, stating that the party “does not claim that it has any monopoly on [Islam] or that it speaks in [Islam’s] name.” To further emphasize this break, the party formally self-identified as a Muslim democratic party. The term “Muslim democrat” has been used by Ennahda for many years, according to Radwan Masmoudi of the Centre for the Study of Islam and Democracy. However, adopting it formally was a deliberate communication strategy to distinguish Ennahda from other Islamist groups ("look beyond
Islamism,” in Said Ferjani’s words. New language also created a new, uniquely Tunisian, political identity. The term “Muslim democrat” intentionally calls to mind the Christian Democratic Union (CDU) party in Germany. Even though the basis of CDU policy is “the Christian understanding of man and his responsibility before God,” because the party welcomes non-Christians, it is accepted as a fully democratic party in Europe. Ennahda’s leadership believes that the party’s relationship to Islam parallels the CDU’s relationship to Christianity.

So why did Ennahda decide to strategically split its religious, civil society movement from its political party? There are several reasons. Some members of Ennahda view the change as primarily pragmatic, with senior Ennahda leader Farida Laabidi saying “Our references will remain Islamic, but it’s not logical for us to try to do everything from tarbiya (religious education) to making economic policy.”

Another reason for this decision is that the goals of the party shifted. Before the 2011 revolution, according to Ghannouchi, MTI/Ennahda promoted and protected Islam in Tunisia because it was threatened by the governments of Bourguiba and Ben Ali, but after the revolution and the establishment of a democratic government, “The state [...] protects Islam. And Muslims and our free society protect Islam. So Islam does not need to be protected by one party.” With the party no longer needing to protect Islam, it could move on to focus on less divisive issues.

Another reason Ennahda may be moving away from political Islam is that politicizing Islam runs the risk of turning Ennahda’s opponents into opponents of Islam. In a country where many people hold a very personalized, apolitical view of Islam, promoting Islam as a political ideology may actually lessen religiosity instead of preserve or increase it. As Said Ferjani told the Hudson Institute just prior to the May 2016 congress, “it’s important to understand that Islamism is one thing, Islam is another,” and that “mixing religious proselytization and party politics] together may harm both.” Ennahda as a political party may be moving away from Islam, but its members do not want Tunisia to have an areligious society. This reason for the shift seems to target the goals of maintaining stability (by seeking to reduce polarization around religion) and winning elections (by opening the party to less religious voters). The latter is particularly evident, especially with Ghannouchi insisting in 2016 that “we don’t ask people to elect us because we are more religious than others. We would like to attract people to our movement regardless of their religiosity.”

Arguably the most important reason for this step away from political Islam, including rebranding to Muslim democrats, is that the move clearly counters the obstacle of association with political Islam. As discussed in Chapter two, Ennahda’s ties to regional Islamism have repeatedly influenced domestic perceptions of Ennahda. In 2013, for example, a coup removed the Muslim Brotherhood’s Mohamed Morsi from his position as Egypt’s president. The move empowered some secularists to call for Ennahda’s removal from power in Tunisia. Ghannouchi decided not to defend Morsi after the Egyptian coup, and subsequently Ennahda’s leaders re-evaluated the party’s relationship to other regional groups that it held ties to, like Qatar, Hamas, and Turkey, and:

“decided that such ties had weakened their own political aspirations by conflating the challenges and controversies of other Islamist organisations with those of their movement. They henceforth distanced themselves from them and claimed to aspire to become a ‘true’ Tunisian party, openly disassociating themselves from the international ambitions entrenched in the Muslim Brotherhood’s ideology.”

A similar case occurred with the AKP, Turkey’s Justice and Development Party. In the 2000s, Ghannouchi and Ennahda openly supported the AKP, especially the party’s gradual approach to Islamizing Turkish society. However, after 2013, the party and its leader Recep Tayyip Erdogan became increasingly repressive and authoritarian. Since this repression began, Ennahda has remained largely silent on the Turkish party, despite its past public praise. In early 2014, Ghannouchi avoided talking about Turkey’s relationship with Islam, painting its domestic problems as primarily rooted in its interventionist foreign policy. By early 2019 he completely refused to talk about Erdogan in an interview with al-Monitor.

Ennahda’s step of moving away from political Islam, emblemized in the adoption of the term “Muslim democrat”, is largely an attempt to separate the party from the discourse on political Islam, which it cannot predict or influence in any substantial way but which influences the party in profound ways. It indicates a symbolic break from regional Islamic political movements, as well as a terminological rejection of the loosely-defined and controversial “political Islam” and “Islamism.” In its party platform created at its 2016
council, Ennahda developed “a new vision of modernity that prevents [the party] from falling in the dilemma of imitation and dependency” on other regional models. Removing itself from contentious discourse is evidenced indirectly with its adoption of the term Muslim democrat and more directly by refusing to engage in discussions on Turkey.

MOVING AWAY FROM POLITICAL ISLAM: SUMMARY
This step of moving away from political Islam is a provocative one. The words “moving away” are chosen intentionally. The term “De-islamization” is not accurate for the efforts that Ennahda is taking, as it maintains Islam as a reference. The phrase “depoliticization of Islam” is useful, but it describes the possible outcomes of the party’s actions more than the direction of the actions. A paper by Cavatorte and Merone argues that actions related to this step constitute an evolution of Islamism – a so-called post-Islamism. This thesis addresses the evolving actions, as opposed to evolving ideology, in agreement with Anne Wolf’s assessment that Ennahda’s move from political Islam to Muslim democracy “did not represent a fundamental transformation in the movement’s ideology.” In other words, it is premature to deem this move as a permanent ideological abandonment of political Islam; rather, it is correct to say that this step is ultimately a political endeavor. By moving away from political Islam, Ennahda seeks to take the wind out of the sails of Nidaa Tounes and other secularist opponents, since those groups gain support when secular Tunisians perceive Ennahda’s religiosity as influencing its policy-making. The move also allows Ennahda to adopt political strategies that present the party as more palatable to secular voters and Western observers, while not appearing to its base to have abandoned Islam in favour of secularism.

C. ADOPTING MODERATE POLITICAL AND GOVERNANCE STRATEGIES
From the moment Ennahda reemerged after the 2011 revolution, critics feared that the party would take over the government, act unilaterally, or take any number of other extreme political actions in the name of Islam. Perhaps the most compelling step the party has taken to quiet these fears is the adoption of moderate and even concessionary political strategies.

COALITIONS
When the votes in the 2011 National Constituent Assembly election were tallied, Ennahda had earned 37 percent. Although this was a sizeable majority, the party decided to form a coalition government with secular parties Ettakatol and CPR. In late 2011, after the election, Ghannouchi told Foreign Policy, “We have declared since before the elections that we would opt for a coalition government, even if al-Nahda achieves an absolute majority, because we don’t want the people to perceive that they have moved from a single party dominant in the political life to another single party dominating the political life.” Many secularists were outraged at the formation of this coalition, commonly known as the Troika, because it was seen as a betrayal by the secular coalition members to work with the Islamists. Publicly, the creation of the coalition was very significant, bolstering Ennahda’s position that it was committed to pluralistic democracy. There were also major decisions made during the coalition’s rule that would have been political crises without a coalition, such as the drafting of the constitution by the National Constituent Assembly. A number of political scandals emerged during this period, such as Ennahda appointing its own partisans to key ministerial positions – Anne Wolf argues that Ennahda’s initial dominance of the political scene was “certainly one of their biggest tactical mistakes.” Despite its challenges, the Troika endured until the ratification of the constitution. The Troika government ultimately resigned in early 2014, and though its importance as a coalition was diminished by Ennahda’s relative might, it was a milestone in Ennahda’s journey towards political moderation.

The Troika government was ultimately less significant than Ennahda’s coalition with Nidaa, its biggest rival. Nidaa Tounes’s membership includes many members of the Ben Ali and Bourguiba regimes, and Ennahda’s conciliatory approach to dealing with such people during the Troika government allowed Nidaa Tounes to gain influence and support. After Mohamed Brahmi’s assassination in July 2013, Nidaa also helped create the National Salvation Front, which was a joint political and civil society group that aimed to push the Troika to resign. Despite the political battles between Nidaa and Ennahda, the tide turned with the 2014 legislative elections. Ennahda “held a celebration” despite losing to Nidaa Tounes, in part because it felt that one election loss did not mean the end of the party’s success. As the 2014 presidential election approached, Ennahda was initially neutral between Nidaa’s Essebsi and his opponent Moncef Mar-
zouki, an ally of Ennahda. Ghannouchi showed his preference when he said “Ennahda should join Nidaa in a coalition government.”202 This led to internal debates within Ennahda, but ultimately a strong rapport between Essebsi and Ghannouchi led to an Ennahda-Nidaa coalition government. It also led Ennahda to lean toward Essebsi (the eventual winner) over Marzouki — Ghannouchi even voted for Essebsi himself in the run-off election.203

The coalition between Nidaa and Ennahda was shaky from the beginning. Nidaa’s primary raison d’être, after all, is to weaken Ennahda’s political influence; it earned tremendous criticism for its decision to partner with Ennahda. In addition, Nidaa’s governance was especially criticized after the attacks in 2015. Twenty-eight resignations from Nidaa by early 2016 gave Ennahda the leading position in parliament again,205 but the coalition remained until September 2018, when Essebsi and Ghannouchi announced the end of their alliance.206 Though the parliamentary coalition still operates under the Chahed government, there is little real cooperation. The coalition was strategic for Ennahda, which preferred being in power to being in the opposition, and although uniting with its chief rival required a number of concessions, it ultimately put more weight behind the party’s assertion that it seeks compromise and unity.

It is also worth noting that, though this holds no binding power, in 2016, when asked by David Hearst if Ennahda would form a coalition if the party won a majority of seats in the upcoming 2019 legislative elections, Ghannouchi said yes.207

STEPPING BACK
Anne Wolf stated that Ennahda’s early dominance in Tunisian politics after 2011 was a tactical error for the party. Since that error, the party has taken more concerted action to avoid overstepping politically and raising fear among secularists. The resignation of the Troika government in early 2014 demonstrated that Ennahda was ultimately willing to cede power in order to achieve stability. The resignation was unpopular among Ennahda’s supporters, who believed Ennahda was forced into the move or that Ghannouchi was weak-willed for agreeing to the plan. Still, the party held together in the wake of the cession of power.

Another chance for Ennahda to step back came with the 2014 legislative elections. Just ten months after the Troika’s resignation, Ennahda was in a difficult position and lost, as discussed previously, to Nidaa Tounes. Despite the intensity of the campaign, though, and the sharp rhetoric from both sides, Ennahda conceded the election before the results were declared.208 This preemptive concession was strategic, indicating respect for the electoral system and allowing a peaceful transition of power.

Perhaps the most compelling case of Ennahda stepping back on the political scene, part of the larger step of adopting moderate political strategies, was the decision to not enter the 2014 presidential race. Despite Nidaa Tounes’s legislative election victory, the secular party was far from unified and Ennahda’s base may have united more strongly in the wake of the election loss. Despite this, the party decided not to nominate a candidate for the race. Worth argues that this was a “relatively easy” decision to make, since the previous year saw the coup of Mohamed Morsi in Egypt, which illustrated to Ennahda that running for the presidency may be overstepping.209 Not running meant choosing between Marzouki and Essebsi, and Ennahda stepped back further by remaining neutral until the final days, and even then only implying an endorsement of Essebsi.

LIBERALIZING
As the party seeks to achieve its goals, it is forced to contend with its own past as an underground movement. Memories of exclusion and repression prevented transparency in and around the party, but it has recently taken thoughtful steps to liberalize its rules to become a modern political party. In a book of conference materials from the May 2016 Ennahda party congress, a section called Organization Vision states that Ennahda seeks “the establishment of a new organizational culture based on the principles of participatory democracy, institutional action, transparency, and accountability.”210 The effects of these efforts are yet undetermined, especially as the culture of the party will need time to shift. The most prominent reforms pertain to recruitment and membership, religion of new members, and gender equity.

As an underground movement, MTI/Ennahda sought to guard itself against infiltration by government-allied individuals.211 In order to do this, it maintained strict recruitment policies. In a process known as tazkiya a prospective party member required the endorsement of two current members.212 The post-revolution political system no longer required such secrecy. The 2016 party congress “transformed” the tazkiya process into “a simple non-binding recommendation,”
lowering the barrier for non-Islamists to join the party. The 2016 party congress brought further change to membership when the term for morals was “stricken entirely from the list of membership requirements.”

This opened up membership even further, making membership itself less demanding on the lifestyles of Ennahda’s members. Previously, new members were compelled to see politics and religion as related, as Ennahda worked to protect and increase the role of Islam in society. After the 2016 party congress, though, the party chose “to function only as a political party in which membership is no longer exclusive to those who share the same political-religious ethos. New members must, in principle, be exclusively interested in political work.”

Ennahda has also liberalized its stance toward the religion of members, in part by opening its membership to non-Muslims. In a case that received international attention, Ennahda nominated a Jewish candidate, Simon Slama, on its Monastir electoral list in Tunisia’s May 2018 municipal elections. As reported by the Times of Israel, Nidaa Tounes’s head of political affairs called the endorsement “a propaganda operation by Ennahda to seem like an open and tolerant party.”

Slama himself argued otherwise, saying Ennahda has “become a moderate party that no longer mixes religion and politics.”

Ennahda’s endorsement of Slama may largely be attributed to the particularities of Monastir. The coastal governorate is not a bastion of Ennahda supporters — it is best known as the home of Bourguiba — and ended up giving Ennahda a smaller proportion of votes than any other governorate in the 2018 elections. Though promoting its image as a more open and progressive party is a well identified strategy, at this point it is not fair to say that the endorsement is merely propaganda. For a party which only two years prior to the 2018 elections opened party membership to non-Muslims, endorsing a non-Muslim for an elected position marked a clear symbolic move and example of the step of adopting moderate political strategies. According to the Washington Institute, Ennahda’s actions in the 2018 elections sent a signal to secularists that “the classic narrative of countering Islamists cannot be effective anymore in Tunisian local politics.”

Besides religion and membership, the final area where Ennahda has adopted moderate strategies is gender equity. As noted in Chapter three, secular Tunisians hold women’s rights to be an important component of Tunisia’s identity, and Ennahda has proclaimed its belief in women’s rights in an effort to emphasize its Tunisian identity. In the wake of the revolution, activist efforts to establish gender parity in the election system led to an election regulation requiring individual candidate lists to alternate vertically between men and women. Ennahda supported this system, and of the forty-nine women elected to the National Constituent Assembly, forty-two were Ennahda candidates. In his 2016 Foreign Affairs article on Muslim democracy, Ghannouchi expressed his support for balanced gender representation in electoral lists. With the work of activists and widespread political support, Tunisia adopted horizontal gender parity, requiring parties to include equal numbers of men and women on lists around the country. Ennahda joined with secular parties in supporting this policy. After the results of the 2018 municipal elections were tallied, 47 percent of local council positions were held by women.

Women in Ennahda ran for a wide range of positions in 2018 – the most prominent example being Souad Abderrahim, a former member of UGTE (Tunisian General Union of Students), who became the first woman mayor of Tunis in July. In an interview with Carnegie Endowment for International Peace, Abderrahim said of Ennahda, “Everyone knows that it is the party that presented a woman as head of the capital, Tunis. All progressive parties — indeed, all parties — proclaim to support gender equality, but in reality they overlooked and criticized women at a critical moment. But Ennahda presented and supported my candidacy for the municipality of Tunis.”

In late 2018, with the political debate over a proposed amendment to inheritance law, Ennahda’s efforts to promote women’s rights came to a halt. The existing law provides daughters with half the inheritance of their brothers, in accordance with the Qur’an (An-Nisa 4:11), and limits inheritance to one-third of a person’s wealth, in accordance with a statement by Prophet Muhammad recorded in Hadith (collection of traditions containing the sayings of Prophet Muhammad) (Bukhari 4:51). President Essebsi proposed an amendment allowing for equal inheritance. Ennahda’s leadership eventually came out against the proposal, arguing that decodifying a Qur’anic practice in a predominantly Muslim country is unconstitutional and un-Islamic, as well as unpopular. Secularists who support the amendment interpreted this stance as revealing Ennahda’s true ideology to be opposed to secularism and women’s rights. After the amendment
was proposed, but before Ennahda made its stance known, Zeineb Ben Yahmen and Sarah Yerkes wrote that “should Ennahda’s leaders oppose the equal inheritance law, the party’s ‘Muslim Democrat’ strategy of normalization, accommodation, and political integration to gain legitimacy on the national and international levels will collapse.”228

Does the decision by Ennahda to oppose the proposed amendment undermine the goals-obstacles-steps framework? Potentially.

After all, the decision reveals a limitation to the secular progressivism of the party. It has raised questions about the sincerity of Ennahda’s Tunisian identity, relationship to Islam, and leftward-trending policy decisions. It has also rekindled fears among secular Tunisian women that they will be oppressed by Islam. In raising these questions, Ennahda’s decision could inhibit the achievement of party goals and even reinvigorate obstacles. However, there is also the possibility that this decision does not undermine the framework. This specific issue, to Ennahda and to many Tunisians, is not merely a question of human rights but is also a question about the Arab-Islamic nature of the state, which is referenced in the constitution. Essebsi’s proposal is highly controversial because it directly contradicts the Qur’an, believed to be a revelation from God. The proposal is also not popular, with 63 percent of Tunisians (including 52 percent of Tunisian women) strongly opposing it, according to a poll conducted in late 2017.229 With the understanding of the proposal’s apparent contradiction of the Qur’an and its lack of popularity in Tunisia, Ennahda’s decision can be seen as a moderate political strategy. The proposal by Essebsi, though it appears to be a similar proposal to other women’s rights efforts supported by Ennahda, can be understood as fundamentally different because of its contradiction of Islamic law and its widespread unpopularity. Supporting Essebsi’s proposal therefore, is not a moderate political strategy and adopting it would be more radical than the framework allows.

After years of supporting more progressive women’s rights policies, Ennahda balking at equal inheritance could seem like either an aberration or a return to normalcy. Perhaps this issue marks the beginning of a change in the framework. More likely, in the author’s estimation, Ennahda is internally reconciling its evolving identity in the face of political opposition, and complicated issues like this bring about complicated responses. Ennahda’s progressive rhetoric has not changed since its decision, but Ghannouchi has generally avoided the topic. For Ennahda, having spent decades fighting for religious freedom in Tunisia, this proposal seems like a test of the country’s Islamic identity; for Essebsi it is part of his platform as a secular progressive. As Ennahda checks its political opposition’s proposal in an effort to preserve its vision of Tunisia, does it undermine its mission to achieve its goals? Its defense of the will of the people and the Qur’an could earn it new supporters and undermine Essebsi’s efforts, already wavering from internal partisan squabbling. Ennahda is adopting moderate political strategies but is still a conservative party. The fact that it is acting like one should not be a surprise despite all the changes the party has made.

SUMMARY

Ennahda’s decisions are not arbitrary but are a deliberate attempt to achieve party goals in the face of significant obstacles. These decisions may seem uncharacteristic for those who view the party as traditionally Islamist. In fact, the party has shown, through a gradual process of change and shifts in branding that it is willing to alter parts of its identity in order to achieve its broader, more strategic goals. By emphasizing its Tunisian roots and identity, the party distinguishes itself from political Islam and reduces skepticism among secular Tunisians. By moving away from political Islam, Ennahda protects Islam from politically-minded opposition, opening itself to support from and participation by non-traditional Ennahda supporters. By adopting moderate political strategies, the party increases trust both domestically and internationally, presenting itself as a party for all conservative Tunisians. These steps are thoughtful and transformative, even if their non-linear nature makes them difficult to track.

Have Ennahda’s steps made any progress in bringing the party closer to its goals? It is a complicated question, but there is some evidence that it has. In the first elections since 2014, the 2018 municipal elections, Ennahda outperformed Nidaa Tounes with 28.6 percent of nationwide votes versus 20.8 percent. However, it is noteworthy that independent lists, or those not registered with a party, received more votes than either party, with 32.2 percent. This is likely a result of dissatisfaction with the ruling party, Nidaa, coupled with an apathy toward perceived political tension and stagnation. Increasing voter turnout may need to become a part of Ennahda’s goal of winning
elections, but it is evidently not yet a major concern. Though Ennahda claimed victory in the elections, the low turnout does not bode well for questions of political stability in the long run.

**FUTURE TRENDS**
The next major election, in late 2019, will be for the country’s legislature and presidency. This will be the true test of Ennahda’s efforts in pursuit of its goal of winning elections, since it is the first major election since Nidaa won in 2014 and Ennahda’s May 2016 party congress. An increasingly relevant factor in understanding the upcoming elections will be the state of Nidaa Tounes as a party. The ‘big-tent’ party is facing factionalism and has shown real signs of splitting. For Ennahda, the possibility of its chief rival crumbling is intriguing, as it may weaken the secular opposition. While there are numerous reasons for the factionalism in Nidaa, it is possible that part of the reason lies in Ennahda’s steps, as each of the steps undermines Nidaa’s talking points against Ennahda. For example, Nidaa and its supporters oppose Ennahda increasing the role of Islam in politics, but Ennahda’s efforts to move away from political Islam weaken that opposition. The future of Nidaa Tounes, no matter what happens, will have significant effects on Ennahda’s goal of winning elections, but also on the goal of stability.

Ennahda will also have to confront any effects of the alienation of Salafism, an “ultraconservative and literalist interpretation of Islam.” Salafis do not comprise a very large group in Tunisia, but they are politically active. As Ennahda moves to attract secular support and distance itself from Islam, it does so to the frustration of this group. This is evidently a sacrifice that Ennahda is willing to make, but it is an interesting development nonetheless. As Ennahda moves away from Salafism instead of working to compromise with them, the Salafis are becoming politically alienated. For the goal of Western support, it has so far been unsuccessful, not materializing in any substantive way. Though tourism may soon increase, it might be some time before Tunisia is able to court any major investment from the West, unless perhaps the Tunisian government is able to stabilise and confront economic issues in a concerted way to make foreign investments more enticing. If Ennahda’s secular counterparts continue to attract investments and loans from Middle Eastern countries while the West balks, Ennahda may be forced to reevaluate its goal and consider looking east for support.

The near future, especially 2019, will be an important year for Ennahda. Beside the election, there is also the question of party leadership. Rached Ghannouchi has effectively led the party since he co-founded it, but in 2019 his second term as President of the party will end. Party rules only allow for two terms, so this would be the end of Ghannouchi’s official presidential role. Will this mean the end of Ennahda? Said Ferjani says no, but it is possible. It opens the possibility that Ghannouchi will run for President of Tunisia, which would mark a significant pivot from 2014, when Ennahda did not submit a candidate for the presidential race. Even if Ghannouchi opts not to run for Tunisia’s or Ennahda’s presidency, he will still play an important role in developing and articulating Ennahda’s strategies. Ennahda without Ghannouchi, whenever it comes, will be in a difficult position.

**CONCLUSION**
The goals-obstacles-steps framework is useful for contextualizing Ennahda’s actions within the broader political and historical context of Tunisia’s post-revolution democratic transition. By allowing for the categorization of actions, the framework reveals patterns of Ennahda’s behavior while allowing one to analyze and possibly even predict larger shifts in strategy. As Tunisia’s political system evolves, the framework will evolve with it as Ennahda contends with new or resurgent challenges and develops shifting goals.

As Ennahda takes action to overcome fear and suspicion among secular Tunisians, it raises questions about the role of Islam in Tunisia. Will Islam ever play a significant role in Tunisian governance? What role does it play in Tunisian culture? Ennahda’s re-emergence after the revolution re-politicized Islam in the minds of many Tunisians; will its efforts to de-politicize Islam increase religiosity or decrease it? Tunisia aims to be a model for democratization, human rights, and freedom in the Arab world, but in a region where Islam plays such a strong role in guiding individual morality and governance, will Tunisia’s de-politicization of the religion make it inimitable for other countries? In 2016, Ennahda’s party congress defined itself as a “modern movement that harmonizes positive modernization on the one hand with rooting culture on the other.” Islam is part of Tunisia’s culture, and so Ennahda seeks to protect it but the party no longer promotes Islam as a feature of politics. As Ennahda works to achieve its goals, its identity is
evolving quickly. This makes it difficult for outsiders to define what Ennahda stands for, while also making it more difficult for party members to understand their party. As it works to gain the trust of Tunisia’s secularists, Ennahda must be careful to not lose the trust of its core constituency. How does Ennahda working to achieve its goals, affect Tunisian democracy, Ennahda’s own identity, the identities of Ennahda’s opponents, and the identities of Islamic political movements around the world? Observers in the Islamic world may be looking to Ennahda and its place in Tunisia as a model for Muslim democracy, but nothing about Ennahda or its place in Tunisia is set in stone. Much has changed in Tunisia since January 2011, and more changes are yet to come.

Endnotes

4 Throughout this thesis, these words will be italicized when they refer to the framework’s categories.
5 Current refers to the period from late 2014 (post-2014 elections) to the end of 2018 (conclusion of research).
18 Wolf, Political Islam in Tunisia.
20 Hamdi, The Politicisation of Islam.
23 Hamdi, The Politicisation of Islam.
31 Monica Marks, “A Response to Sayida Ounissi’s, ‘Ennahda from within: Islamists or “Muslim Democrats”‘,” Brookings Institution, Rethinking Political Islam (June 1, 2016): 8.
32 Tamimi, Rachid Ghannouchi: A Democrat within Islamism.
39 Michael Koplow, “Why Tunisia’s Revolution Is Islamist-Free,” Foreign Policy (blog), January 14,


44 Lynch, “Rached Ghannouchi.”


51 Ibid.

52 Ibid.


58 Lynch, “Rached Ghannouchi.”


60 For example, French is a widely spoken language in the country to this day.


63 Kora Andrieu et al., “Contrasting Notions of History and Collective Memory in Tunisia” (KADEM, Impunity Watch, and University of York, October 2016), 32.

64 Independents, notably, were the most successful group but cannot be called a party as their affiliations are unclear. Some independent lists are informally party-affiliated. This information is based on information gained through informal conversations with Ennahda party members in Tunisia. Ennahda also opened its lists to independent candidates in some cases (Sfaxi, Emir. “Sustaining Democracy.” Carnegie Endowment for International Peace, May 11, 2018, accessed January 25, 2019, https://carnegieendowment.org/sada/76323.)


69 McHale, Mills, and Sullivan.


71 Hearst, “Rached Ghannouchi Q&A.”


73 Kristina Kausch, “‘Foreign Funding’ in Post-Revolution Tunisia,” Foreign Funding in the Arab World: Myth and Reality of a Political Tool (Hivos, Arab Forum for Alternatives, and FRIDE, 2013).


76 Ajroudi, “Saudi Arabia’s Uncertain Investment in Tunisia.”


79 Ibid.

80 Ibid.

81 Ibid, 34-39. NOTE: The MTI became a political entity, but also maintained religious and civil society elements until 2016, which will be discussed in Chapter 3.

82 Ibid, 45.
83 Ibid, 43.
84 Ibid, 53.
86 Bourguiba was in poor health at the time and Ben Ali had doctors deem Bourguiba unfit to rule. The 1987 coup is therefore known as medical coup. (Wolf, _Political Islam in Tunisia_, 65.)
88 Ghannouchi, “From Political Islam to Muslim Democracy.”
89 McCarthy, _Inside Tunisia’s Al-Nahda_, 62.
90 Tamimi, _Rachid Ghannouchi: A Democrat within Islamism_, 70.
91 A security guard was killed at the RCD offices in Bab Souika, Tunis in February 1991.
92 McCarthy, _Inside Tunisia’s Al-Nahda_, 66.
93 Ibid, 76.
94 Wolf, _Political Islam in Tunisia_, 74–75.
95 Cavatorta and Merone, “Moderation through Exclusion? The Journey of the Tunisian Ennahda from Fundamentalist to Conservative Party.”
96 McCarthy, _Inside Tunisia’s Al-Nahda_, 90–94.
97 Wolf, _Political Islam in Tunisia_, 89.
99 Lynch, “Rached Ghannouchi.”
100 “The Dichotomy of Islam in Tunisia.”
101 Shepard, Burgat, and Salvatore, “Islamism.”
102 “The Dichotomy of Islam in Tunisia.”
104 Hamdi, _The Politicisation of Islam_, 17.
105 Wolf, _Political Islam in Tunisia_, 40.
lim-brotherhood-a-terrorist-organization/.
109 Weymouth, “An Interview with Tunisia’s Rachid Ghannouchi, Three Years after the Revolution - The Washington Post.”
111 Wolf, _Political Islam in Tunisia_, 45–51.
112 Jones, “Portrait of Rashid Al-Ghannoushi,” 19.
113 “Salafism is an ultraconservative and literalist interpretation of Islam” from “A Doomed Relationship: Ennahdha and Salafism,” _Islam and Politics in Post-2011 Tunisia_ (Baker Institute, April 27, 2018), 6.
116 International tourism receipts are expenditures by international inbound visitors, including payments to national carriers for international transport
120 Benmelech and Klor, “What Explains the Flow of Foreign Fighters to ISIS?”

165

122 Zelin, “Tunisian Foreign Fighters in Iraq and Syria.”


126 Kirkpatrick, “Opposition in Tunisia Finds Chance for Rebirth.”


130 Ibid.


134 Lynch, “Tunisia’s New Al-Nahda.”


136 Worth, A Rage for Order, 197–98.

137 Nouira, “Obstacles on the Path of Tunisia’s Democratic Transformation.”

138 Rached Ghannouchi, “Islam and Civil Society in Tunisia” (Lecture, August 7, 1994). (Graciously made known to me by Tamara Sonn)


142 Ounissi, “Ennahda from within: Islamists or ‘Muslim Democrats’?”

143 Andrieu et al., “Contrasting Notions of History and Collective Memory in Tunisia,” 27.


147 Hearst, “Rached Ghannouchi Q&A.”


149 Ruedy, Islamism and Secularism in North Africa, 92.

150 “Interview Transcript.”

151 Ghannouchi, “Deficiencies in the Islamic Movement.”

152 “Interview Transcript.”


154 Weymouth, “An Interview with Tunisia’s Rachid Ghannouchi, Three Years after the Revolution - The Washington Post.”
158 Ibid, 139-141.
159 Ibid, 217.
162 Ibid, 200.
163 Ibid, 204.
164 Ibid, 205.
165 Piser, “How Tunisia’s Islamists Embraced Democracy.”
166 Ghannouchi, “Fight ISIS With Democracy.”
167 Ibid.
169 Other revisions probably occurred with the name change from MTI to Ennahda in 1988, with the crackdown by Ben Ali in the early 1990s, and with the 2011 revolution.
171 Ghannouchi, “Secularism and Relation between Religion and the State from the Perspective of the Nahdha Party.”
172 Ibid.
173 Ghannouchi, “Islam and Civil Society in Tunisia.”
174 Ghannouchi, “Secularism and Relation between Religion and the State from the Perspective of the Nahdha Party.”
175 Marks, “How Big Were the Changes Tunisia’s Ennahda Party Just Made at Its National Congress?”
176 Ibid.
177 Ghannouchi, “Secularism and Relation between Religion and the State from the Perspective of the Nahdha Party.”
178 Hearst, “Rached Ghannouchi Q&A.”
182 Ounissi, “Ennahda from within: Islamists or ‘Muslim Democrats’?”
185 Marks, “How Big Were the Changes Tunisia’s Ennahda Party Just Made at Its National Congress?”
186 Hearst, “Rached Ghannouchi Q&A.”
188 Hearst, “Rached Ghannouchi Q&A.”
190 Weymouth, “An Interview with Tunisia’s Rachid Ghannouchi, Three Years after the Revolution - The Washington Post.”
192 Marks, “Tunisia’s Ennahda: Rethinking Islamism in the Context of ISIS and the Egyptian Coup.”
193 Jebnoun, “Tunisia at the Crossroads: An Interview with Sheikh Rachid Al-Ghannouchi.”
196 Cavatorte and Merone, “Post-Islamism, Ideolog-
ical Evolution and ‘La Tunisianité ’ of the Tunisian Islamist Party Al-Nahda.”
197 Wolf, Political Islam in Tunisia, 160.
198 Ibid, 133.
199 Lynch, “Rached Ghannouchi.”
200 Wolf, Political Islam in Tunisia, 134.
201 Ibid, 154.
202 Worth, A Rage for Order, 217.
203 Ibid, 218.
204 Ibid, 219.
205 Wolf, Political Islam in Tunisia, 160.
206 “Tunisia’s Two Sheikhs Are Frenemies No More.”
207 Hearst, “Rached Ghannouchi Q&A.”
208 Ghannouchi, “From Political Islam to Muslim Democracy.”
209 Worth, A Rage for Order, 217.
211 Ibid, 201.
213 Ounissi, “Ennahda from within: Islamists or ‘Muslim Democrats’?,” 4.
214 Marks, “How Big Were the Changes Tunisia’s Ennahda Party Just Made at Its National Congress?”
221 Ghannouchi, “From Political Islam to Muslim Democracy.”
222 “Horizontal parity requires that municipal election lists across Tunisia have equal number of both men and women, while vertical parity requires that men and women alternate within each list.” (UN WOMEN)
226 Grewal, “Can Tunisia Find a Compromise on Equal Inheritance?”
227 Ibid.


230 Ajroudi and Allahoum, “Tunisia’s Nidaa Tounes in Shambles amid Political Turbulence.”


232 Ibid.


Bibliography


———. *Al-Hurriyat Al ‘amma fi Al-Dawla Al-Islamiyya (Civil Liberties in the Islamic State).* Beirut: Center for Arab Unity Studies, 1993.


———. “Deficiencies in the Islamic Movement.” Translated by Linda Jones. *Middle East Report*


Koplow, Michael. “Why Tunisia’s Revolution Is Isla-


Marks, Monica. “A Response to Sayida Ounissi’s, ‘Ennahda from within: Islamists or “Muslim Democrats”?’” *Brookings Institution, Rethinking Political Islam* (June 1, 2016): 8.


PHOTOGRAPHS AS INFLUENCERS: SEARCHING FOR A CONNECTION BETWEEN VISUALS AND SUPPORT FOR LEGAL RIGHTS FOR NATURE

ELZA BOUHASSIRA, SCIENCES PO-COLUMBIA UNIVERSITY CLASS OF 2019

ABSTRACT

In the United States legal system, nonhuman animals and other aspects of nature are treated as property, meaning that animals and land can be bought and sold in a way similar to a table or chair. Some lawyers are working to change this and obtain a form of legal personhood for nonhuman animals and/or ecosystems. One approach, known as Earth jurisprudence, aims to change governance systems so that they consider humans as part of a wider Earth community. The nonprofit the Nonhuman Rights Project takes a second approach and aims to have the rights to bodily liberty and bodily integrity recognized for at least some nonhuman animals. A third approach, called the similarity argument, aims to grant personhood primarily to animals similar to humans. Complex ideas like these can sometimes be represented by imagery; photography can be a powerful tool for communicating complex problems because of the many symbols that can be in a single photograph. This study aims to assess whether photographs can represent these concepts and emphasize the need for a legal structure based in sustainable ideas and which protects nonhuman animals by granting them a form of legal personhood. Eight expert interviews were conducted to understand lawyers’ and photographers’ perspectives on these issues and 289 law students responded to one of three versions of a survey. The different versions consisted of the same set of questions with either no photographs, photographs that matched the questions, or photographs that did not match the questions. Photographs meant to evoke negative emotions resulted in respondents agreeing with statements more strongly than photographs meant to evoke positive emotions. Furthermore, the survey with no photographs evoked more supportive responses than either of the versions that included photographs. The differences in responses were measured using an ANOVA test and a post-hoc Tukey test. The results suggest that law students are more supportive of the theories when presented with either photographs meant to evoke negative emotions or solely-text based information. Understanding how photographs impact responses to these philosophies may reveal which communications strategies could be used to most effectively convey the ideas being studied.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my seminar advisor Suzanne Mac-ey and my research mentor Robin Andersen for their expertise, support, and patience throughout the thesis-writing process. I am very grateful to Suzanne for all of her feedback over the course of the year and to Robin for her help in selecting a project and working through the process of refining the idea as obstacles came up. I would also like thank my second reader, Jenna Lawrence, who provided valuable feedback along the way. Thank you to Jackie Klopp for her help and guidance through the IRB approval process. Thank you to Xiaoyun Qin and Zhenhuan Xie
of the Statistics Department consulting service for their expertise and help designing the methods used for quantitative data analysis. I would also like to thank those interviewed for taking the time to participate in my research. Finally, I would like to thank my friends and family for their continued support through all the different phases of my project.

Watching the Earth rise on December 24, 1968 from the first manned mission to the moon, Command Module Pilot Jim Lovell saw the blue orb of planet Earth contrasted against the blackness of space and said, “the vast loneliness is awe-inspiring and it makes you realize just what you have back there on Earth.” With his statement, Lovell tapped into the stream of thought that encourages humans to act as part of the Earth system by recognizing how dependent we are on our own “spaceship” Earth. There are different iterations of this idea, but they all return to the central principle that human survival is dependent on the Earth’s systems and therefore humans must incorporate that knowledge into their actions.

Considering humans as part of a greater Earth system stands in contrast to the dominant legal philosophy in the United States that sees humans as separate from nature and allows humans to legally exploit it. Photographs may be a way to facilitate the diffusion and discussion of ideas that seek to create fundamental legal rights for nonhuman animals and nature. The example of Earthrise (see Figure 1) is relevant because photographs can impact how people contemplate an issue by evoking emotions and memories. Additionally, the aesthetic beauty of nature captured in a photograph consistently elicits emotional responses in a way few other forms of communication can. Considering the potential impact of photographs and the contrast between these ideas and the US governance system, this study aims to assess the role photographs might play in symbolizing the principles of concepts in favor of creating legal rights for nature.

THEORIES OF FOCUS
Three theories that propose different approaches to create legal rights for nonhuman animals and nature are examined in this paper. They are Earth jurisprudence, the Nonhuman Rights Project (NhRP) approach, and the similarity argument; each is elaborated upon in the coming pages. The NhRP and the similarity argument are descended from animal rights jurisprudence ideas.

Of the three approaches discussed below, this paper will mainly be focusing on the Earth jurisprudence and NhRP perspectives; Earth jurisprudence requires dramatic ideological alteration of the US legal system while the approach taken by the NhRP is a less dramatic change which could bridge the gap between current legal systems and an Earth jurisprudence based system. Furthermore, the NhRP point of view aligns with the goals of Earth jurisprudence to a certain extent in that both seek to expand the bounds of legal personhood. The similarity argument is less closely examined because the other two theories would achieve a similar outcome with potentially fewer drawbacks. It is included because there exists a trend in “the history of social justice activism in the United States [that] suggests that the use of the similarity argument during the initial stages of reform may be inevitable.” Consequently, the similarity argument may play a role in giving nature legal rights—it could act as a stepping stone to one of the two other theories’ goals—and so was included in the expert interviews.

GENERAL OVERVIEW OF STUDY
This study aims to assess whether or not photographs can create associations that evoke a more favorable view of Earth jurisprudence and the NhRP approach so that they can be understood and discussed in legal spheres where lawyers have the ability to enact change using law students enrolled at American-Bar Association (ABA) approved institutions as a sample study group for a survey; they will be referred to as “law students” throughout the paper. The expert opinions of seven lawyers and one photographer are included to gain added perspectives on the two theories tested in the survey and on the similarity argument which
is excluded from the survey. Based on the role that landscape photographs played in the establishment of Yosemite National Park, and the role that other visual icons play in environmental advocacy campaigns like *Earthrise*, it is likely that photographs can also be recognized as influencers on the sample of law students that increase acceptance of ideas which aim to create legal rights for nature.

AN OVERVIEW OF RELEVANT PHOTOGRAPHIC REPRESENTATIONS OF NATURE

A photograph can capture the world in ways the naked eye cannot: a camera can freeze motion at thousandths of a second or collect light for hours at a time. There are two different understandings of what a photograph is. The first says a photograph is a depiction of reality, and the second says that a photograph is a tool that relies on “conventions of representation to be understood and responded to appropriately.” Both of these perspectives are underlain by an innate confidence that the medium of photography shows what really happened when an image was created. The blurring of the line between reality and representation in a photograph means that photography can function as a “structured and structuring space within which the reader deploys and is deployed by what codes they are familiar with” to understand what they are seeing.

In other words, photographs are not just a snapshot of a moment: they are coded representations of an event that happened. When the event was photographed it was embedded with common symbolic modes and influenced by past artistic movements, adding an intellectual background to the image which can change how a viewer sees that moment. These factors are then packaged into a carefully composed and exposed frame that a viewer then experiences as a “photograph.” These qualities make it so that observing a photograph can train the eye to view images a certain way and in doing so can offer new perspectives.

As noted above, the picture of Earth from space, known as *Earthrise*, forever altered our view of our own planet; the Earth was cast as our spaceship. To illustrate another example of how photographs have influenced our views of the Earth and our place within the natural world, take the case of American landscape photographer Carleton Watkins and the modern conception of wilderness. Conceptualizations that define wilderness as sublime in the sense that it evokes intense passion, admiration, and awe while also being powerful and vast, together with the idea that humans live in an unnatural world apart from wilderness, are derived from photographs taken by Watkins. Watkins changed how many Americans saw and thought about wilderness. He created a lens through which American wilderness was seen as a thing apart, a sublime realm of inspiration, untouched by human alterations. By contrast, this human-free “natural” landscape was inherently distinct from the creations of “man,” and became a place of escape from human environments and the social world.

The power of Watkins’ photographs of Yosemite to influence those who viewed them cannot be overstated; they led to the creation of Yosemite National Park and defined the visual conventions of landscape photography for a century. When Watkins’ work was presented to U.S. legislators, they led “Congress to create the world’s first wilderness park,” and set the “trajectory of environmental politics for its first one hundred years.” Watkins began as a survey photographer for the United States Geological Survey and the Army Corps of Engineers. The two organizations employed photographers to help survey new lands in the American West. Survey photographers eventually marketed their work as fine art and founded the field of landscape photography, thereby establishing the precedent for how wilderness was to be framed.

![Figure 2: El Capitan (1861) by Carleton Watkins](image)

Watkins crafted wilderness as something sublime through a variety of techniques, including choosing high vantage points that put the viewer in a position of power, using a large depth of field to enhance the feeling of the hugeness of the landscape, and very rarely including people in his landscape shots. The way of seeing he crafted is so embedded in how we view the world today, that an image of Yosemite’s El Capitan (see Figure 2) is not just viewed as a wall of granite. Rather, it “induces reverence of nature’s vastness” and is associated with escape from society...
and pristine beauty.

Similar representational systems have been established in other areas of environmental communication. For instance, a polar bear on an iceberg is not just a bear living in its habitat; it is a symbolic indicator of climate change, mass extinction, and uncertainty of the future of the planet. Other images of arctic animals and landscapes can also be viewed as similar symbolic indicators. This visual code has been very successfully and widely established over time and such photographs are broadly connected to the representational field where a viewer sees a polar bear and associates the image with climate change.

In addition to the case studies used above, various studies have shown the impact that visuals can have on decision making and opinions. For instance, Dai et al. (2018) showed that in e-commerce, a seller with a profile photo that a buyer considers trustworthy makes more sales than a seller with a less trustworthy profile photo, suggesting that the photographs changed people’s decisions on who to buy from. Feldman and Hart (2018) found that images about actions which can be taken with regard to the climate that are used in news reporting can increase people’s hope. Rickard (2016) found that people who were shown a hurricane forecast in addition to a photograph of a house being flooded were more likely to perceive greater risk and talk about evacuation than those who saw no image, pointing to the idea that photographs reinforced the impending danger in a way words did not. Brascamp, Blake, and Knapen (2015) found that the visual cortex makes some choices between perceptions without a connection to the executive brain areas. In other words, the visual processing parts of the brain can make decisions about what to perceive on their own. Overall, these studies serve to show that visuals can play an important role in decision making, risk perception, and the judgment of an issue.

In summary, photographs are more than depictions of reality; they convey a great deal of meaning based on visual conventions and cultural associations. They carry within them unspoken conceptualizations and emotions that are able to influence how people reflect on a subject in both conscious and unconscious ways. For instance, Earthrise depicted human dependency on the planet and its systems, underscoring the need to protect the Earth. Nevertheless, much of the body of wilderness photography has set the natural world apart from human actions, leaving it to stand as a mythic and inspiring realm. This study aims to see if photographs can evoke positive responses toward the idea of creating legal protections for the natural world based on human interconnectedness to environmental systems and toward the idea of creating legal rights for nonhuman animals: can a viewer see a photograph of a landscape and a person, for instance, and make a link to the interconnectedness of the Earth system and the need to preserve our own habitat? Can a viewer see photographs of animals that underscore their status as autonomous beings and make an association with the need for legal personhood? This study is the first to investigate a possible relationship between photography and the three theories mentioned above.

A BRIEF OVERVIEW OF RELEVANT LEGAL, ETHICAL, AND PHILOSOPHICAL THOUGHT

This section provides an outline of the ethical, philosophical, and legal schools of thought that led to the development of three different approaches to giving nature and/or nonhuman animals legal rights. Understanding the history of these ideas is necessary to best recognize which photographs could capture their core principles. The three theories have different histories, advantages, and flaws, but all aim to ultimately provide more legal protections to the Earth.

Both the Earth jurisprudence and animal rights jurisprudence schools of thought seek to create legal personhood for groups that are currently being harmed by the way our society functions. Ecosystems around the world are suffering due to air, water, soil, light, and noise pollution and their condition will continue to worsen as anthropogenic climate change continues. Animals are suffering due to their treatment in factory farms and their use in experiments, among other causes. The proponents of these theories believe that the various types of harm these two related groups are experiencing could be alleviated if they were recognized as persons with rights that cannot be violated and which can be defended through the legal system.

EARTH JURISPRUDENCE

The first theory is called Earth jurisprudence; it aims to achieve a shift in human governance systems so that people govern themselves as one part of nature rather than treating themselves as separate from nature.

ETHICAL THOUGHT BEHIND EARTH JURISPRUDENCE

American ecologist and philosopher Aldo Leopold wrote about what he called the “land ethic” in the 1940s. The land ethic is the idea that care for the land
Another important piece of the puzzle is understanding Two Treatises of Government, Locke considers private property and how important protecting the rights to include animals, plants, and the land. Leopold argues that existing frameworks need to evolve once again to expand the boundaries of the word community to include animals, plants, and the land.

Essentially, the land ethic is meant to oversee how humans interact with nonhuman elements on the planet and make it so that “a thing is right when it tends to preserve the integrity, stability, and beauty of the biotic community.” In this view, the natural world carries intrinsic value, it is valued in and of itself, and also provides humans with an Earth community that they are wholly dependent upon for survival. The ideas of the land ethic are visible in Earth jurisprudence since both focus on viewing people as one aspect of nature in a greater system.

Deep ecology is a philosophy that builds on Leopold’s land ethic and which also may have played a role in the development of Earth jurisprudence because of its more concrete calls to action. Deep ecology aims to support the intrinsic value of living beings regardless of their role with respect to human agendas. It specifically advocates for reshaping human society in line with that principle. Earth jurisprudence is the application of this school of environmental ethics to the law.

LEGAL THOUGHT BEHIND EARTH JURISPRUDENCE
Another important piece of the puzzle is understanding where the definition of property comes from in Western thought and what is legally considered property today in the United States. Property is an object that belongs to one person or organization. There are different theories as to what led to the emergence of private property and how important protecting the rights of property owners should be in society. For instance, as summarized in O’Brien (1996), in John Locke’s Two Treatises of Government, Locke considers private property an equalizer in an unjust world and deems protecting private property an essential part of a legal system. In his Discourse on Inequality, Jean-Jacques Rousseau sees private property as something that created status inequalities and puts a lesser emphasis on protecting it. Locke had a significant influence on the founding principles of the U.S. government so the American legal system puts a great deal of emphasis on owning property and having a wide berth of what can be done with that property. Currently, all animals, plants, and nature in general are legally considered property and have no rights in the U.S. despite their having legal protections regarding their safety.

It is important to note that legal protections for animals do exist, such as federal and state animal cruelty laws or smaller scale statutes like the banning of declawing of cats in the city of West Hollywood, California. The 1966 Animal Welfare Act (AWA) is a federal statute meant to protect animals in facilities that breed animals commercially, use animals for research, transport animals commercially, or publicly exhibit animals. It establishes minimum care standards for “handling, housing, feeding, watering, sanitation, ventilation, shelter from extremes of weather and temperatures, [and] adequate veterinary care” for animals in those facilities. Its limited scope means that it does not apply to pet stores, farms, or hobby breeders. The act defines the term “animal” as “any live or dead dog, cat, monkey (nonhuman primate mammal), guinea pig, hamster, rabbit, or such other warm-blooded animal, as the Secretary may determine is being used, or is intended for use, for research, testing, experimentation, or exhibition purposes, or as a pet.” It contains exclusions for birds, rats, or mice that will be used in research, livestock that will be used for food or fiber, and for reptiles, amphibians, fish and invertebrates. In other words, the act creates some protections for some animals, but allows exclusions for most of the animals that people use, limiting the actual protection that it provides to the animals in question.

On November 25, 2019, President Trump signed the Preventing Animal Cruelty and Torture Act (PACT), the first U.S. legislation to make acts of animal cruelty punishable with fines and up to seven years’ imprisonment at the federal level. The legislation outlaws acts in which “one or more living non-human mammals, birds, reptiles, or amphibians is purposely crushed, burned, drowned, suffocated, impaled, or otherwise subjected to serious bodily injury.” The bill is meant to cover holes in the existing network of state animal cruelty laws—all 50 states have their own provisions against animal cruelty, but there was no comprehensive law. The new law will help protect animals in cases that cross state borders or which occur in places under federal jurisdiction like airports. Such cases were previously difficult to prosecute due
to differences between state legislation. Despite the victory that the law represents for reducing the suffering of animals, it differs from a rights framework in that it still considers animals’ lives as less valuable than human property and recreation: The bill enumerates exemptions for when it is still acceptable to kill an animal and includes “normal veterinary, agricultural husbandry, or other animal management practice; slaughter of animals for food; hunting, trapping, [and] fishing;” medical or scientific research; and actions that are necessary “to protect the life or property of a person.” In essence, the law creates stronger protections for animals, but still considers them objects and not rights holders.

Regardless of the greater focus that has been placed on protections in legislation and court rulings, there have been some instances where the rights of nature have been mentioned in U.S. law. In the 1970s, the United States saw a court case which addressed the question of the rights of nature and it also enacted further legislation dedicated to managing the nonhuman animals within its borders. In 1972, Sierra Club v Morton, a case looking at whether the Sierra Club had standing to sue to protect a valley from being developed, was dismissed on the grounds that the Sierra Club had not proved the valley would be sufficiently harmed by the proposed development to have grounds to sue on. The case was notable partly because of the dissenting opinion of Justice William O. Douglas which argued that existing doctrine in the U.S. should give environmental organizations like the Sierra Club the ability to bring lawsuits on behalf of inanimate objects like a valley. His dissent made an argument similar to the one that Earth jurisprudence makes in the sense that it sought to allow lawsuits to be brought on behalf of nature, though it did not have concrete impacts. There have been legal cases in the U.S. like Scenic Hudson Preservation Conference v FPC in which some latitude was given to hearing a petition on behalf of nature, in Scenic Hudson the case concerned a mountain.

Since these cases were decided, there have been new developments. For instance, on February 26, 2019, residents of Toledo, Ohio voted to pass the Lake Erie Bill of Rights (LEBOR), a document which gives Lake Erie and its watershed the “right to exist, flourish, and naturally evolve.” The bill is one of the first of its kind to be adopted in the U.S. and aims to make it so that Lake Erie would not need to show injury to a human in order to sue for wrongdoing. LEBOR was quickly challenged: on February 27, 2019 a lawsuit was filed questioning its constitutionality. Drewes Farms Partnership v City of Toledo, Ohio claims that LEBOR is a governmental overreach, is unconstitutional, and stands to harm local farms and businesses. LEBOR was a step forward in the US for Earth jurisprudence, but has a long road ahead as it goes through the lawsuit which was filed.

In terms of legislation, the U.S. Fish and Wildlife Service (FWS) initially approached management of wildlife in the U.S. by focusing on consumptive activities like sport fishing and hunting. A shift occurred in the U.S. legislative approach to the environment in the 1970s with the passing of the National Environmental Policy Act of 1969, the Marine Mammal Protection Act of 1972, the Endangered Species Act of 1973, and the Convention of International Trade on Endangered Species of 1973, which created a focus on protecting biodiversity, endangered species, and more generally, ecosystems and the environment from human impact. Nevertheless, all of these acts came after the 1964 Wilderness Act which legally defines wilderness as “an area where the Earth and its community of life are untrammeled by man, where man himself is a visitor who does not remain.” Consequently, there still exists a categorical difference between people and nature.

Legal protections, like those passed in the U.S., differ from legal personhood. Protection laws diverge from giving nonhuman animals and nature a form of legal personhood with the right to exist or to be free from imprisonment because they operate under the current framework that treats animals as property. Legal personhood is the status which gives a natural person (a human being) or an artificial person (a corporation) legal rights, protections, responsibilities, and legal liability. Christopher Stone’s landmark essay “Should Trees Have Standing?” lays out one of the first Western notions of giving nature legal rights. Stone suggests that the legal problems of “natural objects should be handled like those of humans who have become vegetables.” He argues for a guardianship approach to governing nature where an organization (private or public) or an individual could apply for guardianship of a natural entity to protect its interests since it is not able to do so independently. This would give a natural entity a form of legal personhood. One exception does exist: there is a tree which was deeded to itself in the 1820s and as a result, the tree owns the land eight feet around where it sits on all sides, so it
Environmental attorney Cormac Cullinan takes the argument for giving nature legal rights to its extreme. He coined the term “wild law” which refers to law that enacts Earth jurisprudence ideas and encourages a shift to a system based on cooperation with rather than exploitation of the natural world. This study uses the term Earth jurisprudence to refer to Cullinan’s school of thought.

Cullinan believes that if the goal is to protect the environment, the legal system needs to be founded on a land ethic type of worldview. In addition to Western philosophical ideas, Cullinan draws on African customary law and Indigenous groups’ governing codes that contain views of the world that are founded on a deep connection between humans and the Earth.

In the 2010 Universal Declaration of the Rights of Mother Earth, Cullinan elaborates on what specific rights Earth jurisprudence aims to obtain for the natural world. The declaration includes the right to life, to exist, to be respected, to be free from human disruptions, to water, to clean air, to be free from pollution, to restoration for the violation of rights by humans, and to live free from torture or cruel treatment. Cullinan’s approach differs from that taken in the Animal Welfare Act and the Prevention of Animal Cruelty Act because Cullinan includes a wider range of beings, he includes ecosystems as a whole rather than certain enumerated nonhuman animals, and aims to reform governance systems rather than functioning within them as the AWA and PACT do. Cullinan repeatedly quotes protestors from the 2009 Copenhagen Conference of the Parties meeting who chanted “system change, not climate change.” The chants from Copenhagen aptly capture the essence of the magnitude of change for which Cullinan is calling.

Support for Earth jurisprudence is gradually gathering momentum around the world. For instance, Ecuador’s 2008 constitution contains a clause that pledges its citizens and its government to a “harmonious existence with nature.” The topic has even been addressed at the United Nations; in 2009, the Bolivian president gave a speech in which he expressed hopes for the 21st century to be the century of Earth rights. In September 2010, the Global Alliance for the Rights of Nature was founded. It is a network of organizations and individuals committed to the universal adoption and implementation of legal systems that recognize, respect and enforce rights of nature which is still active today with projects unfolding all over the world. Around the world, Earth jurisprudence is trying to fundamentally dismantle the idea that humans are different from other living beings, to get people to stop holding themselves above other forms of life, and to do so by regulating humans as mammals rather than as unique beings.

**ANIMAL RIGHTS JURISPRUDENCE**

The second and third approaches that will be looked at both fall under the umbrella of animal law. The term animal law applies to legal issues that involve animals and as such can be statutory or case law and involve issues that draw in many other areas of law. These ideas differ from those of Earth jurisprudence because they do not include plants or the physical environment. They also come from a different tradition of thought—animal rights thought is descended from animal protection ideas and action.

**PHILOSOPHICAL HISTORY OF THE MODERN ANIMAL RIGHTS MOVEMENT**

Prior to the early 1970s, most discussion regarding the treatment of animals was centered on animal protection; actions that are meant to protect animals from undue harm, but which still treat them as property under the law. Books like Upton Sinclair’s 1906 *The Jungle* and Ruth Harrison’s 1964 *Animal Machines* criticized factory farming practices, brought public awareness to the mistreatment and exploitation of animals in factory farms, and fostered thinking about the treatment of animals.

A turning point in the animal protection movement was reached at Oxford University in the early 1970s when a small group of students led by Rosalind and Stanley Godlovitch, students at the university, founded the Oxford Group to discuss the idea of animal rights among themselves and with the philosophy professors at Oxford. British psychologist Richard Ryder joined the group and wrote a pamphlet to be given out to spread awareness of the ideas they were looking at. In that pamphlet, written in 1970, he coined the term “speciesism.” Speciesism, discrimination on the basis of which species an organism belongs to, is when one organism is given better treatment than another in a matter in which their interests are the same. For example, a cow is given less consideration than a dog on the basis of species; it is commonplace to wear leather clothing, but an item made from dog skin would not be seen as normal by most people. Yet another member of the Oxford Group, philosopher Peter
Singer, popularized the term speciesism with his 1975 book *Animal Liberation*. Singer makes the case that animals deserve equal consideration—not equal treatment—on the basis of their ability to feel suffering. He does not, however, believe that rights are necessary for them to receive such consideration; in his view their capacity for suffering alone is enough to warrant the end of their exploitation.

Written in 1983, *The Case for Animal Rights* by Tom Regan shifted the discussion toward a rights framework when Regan made the following case. He claims that since we ascribe inherent value and respect to infants and the mentally impaired, the characteristic which makes all humans deserving of moral consideration is not rationality, but that all humans have a life that matters to them. In essence, even if no one else is impacted by what happens to us, we still care about what happens to us. He calls this being “the experiencing subject of a life, a conscious creature having an individual welfare that has importance to us whatever our usefulness to others” and states that any subject of a life has inherent value and moral rights. A moral right is a right accorded under an ethical system—not moral rights which built on Regan’s ideas was the 1995 *Animal Liberation* by Gary Francione. Francione argues against animal protection and welfare laws in favor of creating legal rights for animals. He makes the case that in the American legal system where animals are considered property, “it becomes an irrefutable proposition that animals lose out whenever humans want to exploit them.” In other words, he sees the system as designed to favor humans’ interests and marginalize those of animals, resulting in a legal system which cannot truly ensure the end of animal exploitation. Francione was one of the first intellectuals in the animal rights movement to explicitly articulate that the legal system, as it stands, will not lead to the liberation of animals. He argues that changes, through something like rights theory, may be needed to ensure that animals receive the treatment he believes they deserve.

Other branches of thought in the animal liberation field sprouted between the works highlighted here—different ideas grew in alternate directions because they were rooted in diverse understandings of which characteristics qualify animals as deserving of moral or legal consideration or rights, how far those considerations or rights should go, and what methods should be employed to bring about change. The overview of the philosophical origins of the modern animal rights movement provided here is by no means comprehensive, but should provide sufficient context of how the movement unfolded to better understand where the second and third approaches addressed in this study come from and how they are different from the first theory.

**THE NONHUMAN RIGHTS PROJECT APPROACH**

The Nonhuman Rights Project approach is the second theory being examined; it continues down the line drawn by scholars like Regan and Francione. The NhRP is a nonprofit founded by attorney Steven Wise in 2007. It categorizes itself as the only civil rights organization working toward rights for nonhuman animals, not an animal rights organization. The NhRP works to obtain the right to bodily liberty and bodily integrity for at least some nonhuman animals like great apes, dolphins, elephants, and whales. Bodily liberty is defined on the NhRP website as the right to be free from imprisonment, and bodily integrity is defined as the right to be free from experimentation. Wise (2000) “demands legal personhood” for chimps and bonobos and an end to biomedical testing, training, and the practice of keeping these nonhuman animals as pets. The NhRP looks at these nonhuman animals specifically because “they are members of species for whom there is robust, abundant scientific evidence of self-awareness and autonomy.” It is willing to expand the animals it represents if scientific literature on others species that meet the criteria it considers is published.

The NhRP takes the stance that all entities are both infinitely similar and infinitely different, so it does not argue for rights on the basis of similarity to humans. It instead argues that autonomy is the necessary characteristic for non-comparative fundamental rights. In an interview, Wise described autonomy as the ability to act freely and to choose how to live one’s life; he also mentioned that there has yet to be a formal definition of the term from a judge or legal document.

The organization’s strategy is to bring lawsuits on behalf of individual nonhuman animals with the goal of changing common law by getting a writ of habeas corpus which would recognize that the given
nonhuman animal has certain rights that had been violated. Habeas corpus is a legal recourse through which a person can report that they have been unlawfully detained or imprisoned and request that a court order those detaining the person to determine if the detention is lawful.

The NhRP’s approach is innovative because nonhuman animals have been considered property since the earliest Sumerian law and that classification can be traced through Hebrew, Greek and Roman law. The early origins of classifying nonhuman animals as property means that all Western legal systems have the idea embedded in many of their statutes since they emerged from the same root. However, in the recent past, some countries have recognized nonhuman animals as having some legal rights. In 1992, Switzerland became the first country in the world to amend its constitution to recognize nonhuman animals. In 1999, New Zealand “banned the use of non-human hominids in research, testing, and teaching except where such uses are in the hominids’ best interests.” The NhRP aims to bring about the kind of change occurring internationally to the United States through the U.S. court system.

The first lawsuit the NhRP brought, Nonhuman Rights Project, Inc. on Behalf of Tommy v. Patrick C. Lavery Et Al., was on behalf of a chimpanzee. The NhRP filed a petition, a legal document which states what a case hopes to accomplish, on behalf of the chimpanzee demanding that he be released from unlawful detention. Although the case did not succeed in getting the chimpanzee released to a sanctuary, the judge who ruled on the case did state that the issue “speaks to our relationship with all the life around us” and that “ultimately, we will not be able to ignore” the issue. Since that case, the NhRP has continued to file similar cases hoping to obtain a writ of habeas corpus for an imprisoned nonhuman animal that would recognize that nonhuman animals have certain legal rights and are not just property. The approach taken by the NhRP requires a great deal of persistence. They have already begun five cases, and it may take a long time before such a writ is granted due to its departure from current understandings of who can be granted a writ of habeas corpus.

THE SIMILARITY ARGUMENT
The third point of view being examined is the similarity argument. It is the idea that some nonhuman animals are so similar to people they should have rights due to their similarity. In other words, at a certain point some nonhuman animals share enough characteristics that are used to define humans (i.e., use of tools, complex social relationships) that they should have comparable fundamental legal rights to those humans possess.

The similarity argument is highlighted in this study because both the NhRP approach and Earth jurisprudence seek to expand the bounds of legal personhood, and it could potentially act as a stepping stone to creating rights for dissimilar animals or physical nature. It calls for a less dramatic change to how the American legal system functions than would Earth jurisprudence, for instance. Earth jurisprudence calls for a fundamental shift in how humans and the legal system relate to the natural world while the similarity argument continues trends that were started by past rights movements. Likewise, “civil rights activists, feminists, gay rights advocates, and the disability rights community have all used the argument that” their respective populations deserved rights because of their similarity to groups with legal rights. If the animal rights movement unfolds similarly to any of the past movements to create rights for a population, it is possible that it will follow a parallel trajectory.

This argument, however, plays a less central role in this study because of its flaws and the possibility that the other two approaches would effect a similar end result with fewer obvious drawbacks to the populations they are trying to protect. Bryant (2007) highlights flaws of the similarity argument. A major one is that it creates a hierarchy among nonhuman animals. By ranking nonhuman animals in order of similarity to humans it places the most value on those that can be claimed to be like people. The hierarchy it creates enables the exploitation of dissimilar animals and could benefit some animals at the expense of others. For example, if sea lions were given personhood and some entitlements on the basis of similarity, certain fish populations would likely suffer so that all sea lions would have enough food to be healthy. Similarly, focusing on the qualities of nonhuman animals that make them like people takes away from the incredible diversity on Earth by only placing value on that which is similar to humans.

SUBSECTION SUMMARY
In summary, the essential message of Earth jurisprudence is that governance systems must be changed to foster a connection between humans and nature
rather than facilitating human exploitation of and dominance over nature. It aims to create fundamental rights for all nonhuman aspects of ecosystems. The second and third theories both come from animal rights jurisprudence schools of thought. The second theory, the NhRP approach, aims to have some rights recognized for some nonhuman animals and plans to do so by having a writ of habeas corpus granted for an individual for whom the organization files a case. Photographs that evoke support for Earth jurisprudence or the NhRP need to capture the core philosophies of these ideas. The specific types of photographs chosen to represent the two philosophies are elaborated on at length in the section on survey methods. The third theory, the similarity argument, aims to create legal rights for nonhuman animals which are similar to humans.

THE INTERVIEWS
APPROACHES AND METHODS
A mixed-methods approach was used and both qualitative and quantitative data were collected. Quantitative data collection was done through a digital survey. Qualitative data collection was done through interviews with experts in the legal arena and photography. The survey as it appeared to respondents can be found in Appendix I. Qualitative data were also collected through interviews with lawyers and a photographer.

INTERVIEWS
One photographer and seven lawyers were interviewed to gain more perspectives on the philosophies being investigated in this study and to determine what some current lawyers working on environmental or animal law issues believe with regard to these theories. Most of the interviewees are not experts in earth jurisprudence or animal rights law. The photographer was interviewed to gain another perspective on the difference between upsetting and beautiful photography of animals and nature. Since the interviewees were experts who can be reasonably expected to comment on the questions posed, interview questions did not need Institutional Review Board approval because interviewees were not placed in a vulnerable position.

Initially, only lawyers were contacted for interviews to gain the perspective of a small sample of practicing lawyers in environmental or animal law. These experts were contacted because they work in different areas of environmental law and have a diversity of experience that would provide dissimilar points of view from within the scope of environmental law. The interviewees work in government, the nonprofit sector, academia, and the private sector. One of the interviewees worked with photojournalist Jo-Anne McArthur and suggested her as a potential interviewee, and so she was contacted to gain an additional perspective.

The guiding script used during the semi-structured interviews can be viewed in Appendix V. Since the interviews were semi-structured, the script was merely a jumping-off point. Sometimes the conversation wandered based on interviewees’ responses and unique expertise. Interviews were conducted over a four-week period which overlapped with the time the survey was open to responses. Table 3 lists interviewees’ names and backgrounds.

<table>
<thead>
<tr>
<th>Name</th>
<th>Background Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jo-Anne McArthur</td>
<td>McArthur is a photojournalist and the founder of the 'We Animals' project, a photography archive that looks at the relationships between humans and animals.</td>
</tr>
<tr>
<td>David Ebert</td>
<td>Ebert is the co-founder of the Animal Defense Partnership, a pro bono legal and professional service provider to the animal protection community.</td>
</tr>
<tr>
<td>Paul Gallopp</td>
<td>Gallopp is the president of Riverkeeper, an environmental nonprofit dedicated to the protection of the Hudson River.</td>
</tr>
<tr>
<td>Michael Gerrard</td>
<td>Gerrard is the Andrew Sabin Professor of Professional Practice at Columbia Law School and has practiced environmental law for almost thirty years.</td>
</tr>
<tr>
<td>Phillip Karmel</td>
<td>Karmel is an attorney and partner at Bryan Cave Leighton Paisner LLP. His work concerns issues of environmental, land use, real estate, and toxic tort issues.</td>
</tr>
<tr>
<td>Justin Marceau</td>
<td>Marceau is a professor and the Animal Legal Defense Fund Professor of Law at University of Denver Sturm College of Law. Currently, his research looks at animal and criminal law.</td>
</tr>
<tr>
<td>Joseph Siegel</td>
<td>Siegel is a Senior Attorney and Environmental Collaboration and Conflict Resolution Specialist with the U.S. Environmental Protection Agency, Region 2, Office of Regional Counsel in New York. His focus is on climate change and air pollution law and policy.</td>
</tr>
<tr>
<td>Steven Wise</td>
<td>Wise is the founder and president of the Nonhuman Rights Project and has practiced animal protection law for thirty years.</td>
</tr>
</tbody>
</table>

Table 3: Interviewee names and backgrounds

RESULTS
BREAKING DOWN THE INTERVIEWS
This section summarizes relevant responses from phone interviews based on recordings of the conversations that were conducted with the eight individuals mentioned in the table. The interviews are broken
down by question with the most pertinent responses listed below. Responses have been edited for length and clarity.

**Lawyers**

**Were you previously familiar with either of these strains of thought? Have you done any work related to either?**

**Philip Karmel:** Karmel was not previously familiar with these theories. After a brief explanation of the NhRP approach, he said, “I don’t think they’re likely to have much success. In general, rights in the United States are grounded in either the Constitution or statutes enacted by the legislature or regulations enacted by an administrative agency, pursuant to a statute. In general, just because someone thinks something is bad does not give rise to a legal claim. You can ground claims in the common law, for example, negligence or something like that. But I don’t think the common law protects animals either. I don’t think that bypassing the legislature and going to the court system is democratic or even likely to be successful at all.”

**Michael Gerrard:** “Somewhat. In my environmental law class, I teach students about Christopher Stone’s “Should Trees Have Standing?” and the dissenting opinion in Sierra Club v. Morton.”

**Joseph Siegel:** “I would say yes. I teach and I’m familiar with the concepts, but I don’t know the subtleties. But I have taught a number of things related to the rights of the Earth, and animal rights comes up a lot in my classes. My students often do independent projects related to it.”

**In your opinion, is giving nature and nonhuman animals some legal rights an effective way to address environmental issues? If not, what legal avenue do you think is more effective?**

**Michael Gerrard:** “The first question is the political feasibility of it; I don’t think it is in any way feasible to do that unless you define it really narrowly. There are a wide variety of legal techniques that can be used—whether through legislation or the courts—which exist as alternatives.”

**David Ebert:** “I think it’s effective in that it makes people more aware of the issues involved,” Ebert said. He expressed doubt in how far those rights could go, giving the example that if chimpanzees were released from confinement, “They have to go somewhere and there are sanctuaries for them to go to, but not for all of them.” He believes that the best way to effect change is to make people aware of the harm occurring because most people are not aware of the atrocities being committed against animals, particularly animals that are used for food.

**Philip Karmel:** “I think social change begins with education, and I think of the tremendous harm that our heavily industrialized, fossil fuel-intensive society does to the ecology of the planet, including the animals who live on the planet,” Karmel said. He continued, stating, “If you’re trying to educate the public, I think you need to use various methods because some people will be influenced by words, some people will be influenced by speech, and some people will be influenced by pictures. Take the most extreme case, someone who is illiterate: If you give them a book, it will have absolutely no effect on their knowledge base whatsoever. The book is meaningless.”

Karmel brought up that there are existing statutes that protect animals and the environment, but he does not think the best path to further protecting them is to create legal rights for them. Karmel believes that most people today would not say that cows and chickens have the right to life in the same way humans do, but that many people, including those in the United States Congress, do believe that species as a whole have the right to exist or at a minimum, that their existence should be protected. An example of this is the Endangered Species Act, which exists to protect species from extinction. He used the example of habitat conservation plans, which are required under the Endangered Species Act to show how the current system works to protect more than just specific species. He believes that “the main reason why animals are on the list of endangered species is their habitat is being systematically intruded upon by residential development, commercial development, industrial development, pollution of some kind through pesticides/herbicides, land clearing, or agricultural activities.”

**Paul Gallay:** Gallay thinks, “It [granting nature and nonhuman animals legal rights] is going to become increasingly more effective.” Gallay cited the 2019 referendum in Toledo, Ohio, to protect Lake Erie as an example of how there has recently been more traction for such ideas in the United States. “Change comes when you actively promote it; most of it until 1965 came from environmental organizations, like Riverkeeper or NRDC. Every movement has to start somewhere.”

**Justin Marceau:** “I don’t think it’s the only avenue and in law, it’s very difficult to measure. I’m a believ-
er in taking multiple paths, and you’re never going
be able to predict exactly what works or what works
best,” he said. Marceau believes that affording rights
“would be something.”

Joseph Siegel: “I think it is an effective way of pro-
moting [environmental rights]. I think we have to rec-
ognize the rights of animals and non-animal species,
and the law is a very important way to get that done.
But I think an equally effective way is through educa-
tion and teaching environmentalism and sensitivity to
other beings in the world to our kids. It’s really neces-
sary to do that as well. You have to do both.”

Steven Wise: “I actually don’t work in environmental
law, I work in civil rights law. The reason that I and
the Nonhuman Rights Project are working to give
nonhuman animals legal rights is for the same reason
that other civil rights organizations would be interest-
ed in human rights. Because without rights, you have
no capacity for rights and you are essentially invis-
ible,” he said. He is against working inside current
systems because doing so treats nonhuman animals as
things, and he believes that is a serious problem.

Another lawyer I interviewed earlier in this project
was against using the court system to change the
status of animals. I think he even used the word
undemocratic. How would you respond to that
point of view?

Steven Wise: “Rights aren’t democratic. My funda-
mental rights are not democratic. I have them. In fact,
you ought to have them even when everybody else in
the world wants to kill or imprison someone arbitrar-
ily. That’s the point of a fundamental right, it’s kind
of like a spoke in the wheel. I think that they stop the
majority from oppressing the minority. The lawyer
may not realize it, but he may be in the majority. His-

torically, if you allow fundamental rights to be subject
to majority rule, then the majority is going to oppress
the minority. That’s why the courts are where, for
example, people of color or women get to assert their
fundamental rights to keep themselves from being
oppressed by others. A fundamental right is fundamen-
tally undemocratic. Because if not, the majority will
go back to oppressing the minority. If it were possible
to get a law through the legislature, we would do that,
but it is not.”

Can you see Earth jurisprudence ideas impacting
the work you do on a day-to-day basis?
Justin Marceau: “In reality, no, not every day, be-
cause not every day are we fighting about the rights of
nature. Sometimes we are fighting to just have a right
to show images. I spent several years just working
on bringing transparency to factory farms. So in that
sense, we’re not really working in the trenches at all
about articulating rights. On the day-to-day level, it’s
not really necessary to be in that kind of conceptual
space, but I always try to keep in mind how the work
would impact, harm, or benefit humans and nature.”

Joseph Siegel: “Yeah, they do. A lot of the work I do
involves the Federal environmental statutes. In partic-
ular, I’ve worked a lot on the Clean Air Act, and there
are a lot of references in the statutes to human health
and the environment. The work that I do derives from
laws that seek to protect natural objects or other spe-
cies.”

What do you think the biggest challenge is facing
the Nonhuman Rights Project?
Steven Wise: “The biggest challenge that we have
is that nonhuman animals have never had rights
and judges have never been taught that they should.
Judges, just like everyone else, have an implicit bias
against the idea. It’s part of the culture in which they
grew up and just like other kinds of bias, like racial
bias or religious bias. The idea is against the culture
they grew up in, so it’s difficult. Judges are human.
For judges to rule in our favor, the first step is to un-
derstand where their biases are. Although some may
be biased in our favor, generally they are not. That’s
our biggest challenge.”

The third argument that is being looked at is that
nonhuman animals which are most similar to hu-
mans deserve some legal rights more than nonhu-
man animals which are dissimilar. Do you believe
the similarity approach is a necessary first step
toward a greater shift in the system, if one were to
occur?
David Ebert: Ebert thinks that there needs to be a dis-
tinction between which nonhuman animals would be
granted rights and which would not because it would
be impossible to grant all nonhuman animals legal
rights. “It is easiest to convince people that animals
with the ability to make choices and an awareness of
themselves as independent beings should have rights,”
he said. He emphasized that the aspect he personally
has the most concern about is alleviating the suffering
of animals, not whether more or less similar animals
are more or less deserving of rights. “It truly keeps
me up at night. It’s going on right now. I went after a
processing plant in Minnesota a couple of years ago and they process 19,000 pigs a day. 19,000 animals go through this each day in just horrific conditions. That’s a lot, a lot, a lot of suffering and that’s where my interest, as I said, keeps drawing back to that issue as the central issue right now that has to be dealt with,” he said.

**Paul Gallay:** “It could be a good way to start, but I’m a bigger believer in the idea of rights to systems as opposed to the more recognizable denizens of those systems. There’s a lot of interesting research being done right now about how you get people to connect with nature and not see themselves as apart, but see themselves as embedded in nature. I think if you’re going to have embedded nature, you should be part of a system more than thinking of yourself in relation to the charismatic megafauna—just a personal opinion.”

**Michael Gerrard:** “Yes. I think there would be more sympathy toward giving rights to chimpanzees than to wasps. I think the more cognate the creature, the more I would be concerned about treating it cruelly.”

**Philip Karmel:** “Yes. I think that it’s much more likely to be successful to take a selective approach. Say I’m trying to protect a certain animal rather than every animal, so the dog, an elephant, a sapien, monkeys—you know, things like that. I think that’s more likely to be successful because if you tried to pass a law prohibiting the extermination of cockroaches in apartment buildings, no one would support that whatsoever. It’s inconceivable to me that any legislature or court would support that. But obviously, an elephant is considerably more advanced [than] a cockroach, and elephants have well documented emotions, intelligence, family structures, communities, and social organization. So I do think that’s a distinction that is more likely to be successful than some blanket assertion that all animals have rights.”

**Justin Marceau:** “I think people believe that if you start with dogs and cats or you start with the chimpanzees and elephants, they could act like a gateway drug that leads to greater awareness in the other spheres. I think it’s possible; I don’t have all the inputs. I think it’s also dangerous to the extent that we are diminishing the suffering of pigs or cows to prop up chimpanzees. We have lots of experts talking about the cognition of whales and elephants and primates. We gauge everything else’s worth by how it’s like us. But dolphins can speak their own language that we can’t even understand and speak in frequencies that we don’t even pretend to be able to hear, so [I] mean I think there’s a danger, but I also understand why one would choose to start with the most charismatic and loved animals.”

**Joseph Siegel:** Siegel believes that as a practical matter the best way to approach creating rights for nonhuman animals and nature is to use the similarity argument because some animals resonate more with people, and he is in favor of using whatever strategy is most effective for getting people to be more aware of environmental issues. “Ultimately, choices will be made: Are we going to give the same rights to a cockroach as we would to a bear? Those get to some very difficult kinds of questions. But from a strategy perspective, I think that a lot of folks in the climate change world have focused on the polar bear because it’s sympathetic, but there’s a lot of other species in the Arctic that are being affected by climate change. But a polar bear just resonates with people, and I’m okay with using whatever works to raise people’s awareness,” he stated.

**Steven Wise:** “The NhRP does not believe similarity is the characteristic we should consider,” Wise said. Instead, it considers autonomy. The NhRP recognizes that giving legal rights to a nonhuman animal will sound strange and raise problems for judges at first because it has never been the case that nonhuman animals have rights. To minimize the difficulty judges have in accepting this new idea, the NhRP studied “the principles and values that judges within a jurisdiction say that they believe in and the evidence that [the NhRP has] been given to see what principles and values judges say they believe in” are also in the facts of the case. The NhRP has found that “judges greatly value the idea of autonomy.” Wise said that the NhRP is in this fight for the long run and is prepared to convince judges over time as they see the argument “five times, ten times, twenty times [because] the first time they see, it it’s unlikely that they’re going to really grasp” what the NhRP is doing.

**Could you define autonomy?**

**Steven Wise:** “That’s a good question because the judges don’t really define autonomy. We bring in experts on autonomy, but it seems that experts kind of agree implicitly with what the judges are saying. Autonomy is the ability to act freely, to choose how to live one’s life. If you do something under duress, you lose your autonomy.”

**Do you think it is possible that the idea to give**
nature and nonhuman animals some legal rights could become mainstream in legal scholarship?
Michael Gerrard: “Well, more mainstream. I mean maybe instead of appealing to 0.1 percent of the population, [it] will appeal to 0.2 percent of the population.”

Philip Karmel: “I think strengthening wildlife protection laws could become more mainstream. Going to court and trying to create new rights for animals that were never enacted by Congress or state legislature, I don’t think that’s going to work and I don’t think that’s going to become mainstream. In my lifetime, many laws have been enacted to protect animals including at the Federal level: the Endangered Species Act, the Magnuson–Stevens Fishery Act, the Bald Eagle and Golden Eagle Protection Act, the Migratory Bird Treaty. We pass a number of laws to protect wildlife, and I could see enacting more laws. But we’re not going to just create rights in the court system. I don’t think that’s going to work.”

Paul Gallay: “I think it can get some traction. I think that a lot of ideas about the environment are shifting right now because of things like the Paradise Fire and the Polar Vortices we’ve been seeing. It just seems that people are getting very anxious about their ability to exist outside nature. I think as a consequence, they’re going to be looking for ways to make peace with nature and, for lack of a better term, to join forces with nature.”

David Ebert: “Yeah, I do. There were a couple of judges who were at least raising the possibility that this is something that we should be looking at in decisions on Steven Wise’s cases. I think that it certainly is a body of law that could be developed, that could be debated, and you could make laws that make sense and then make a difference. Yeah, I think it can be part of the mainstream.”

Justin Marceau: “I think it will balloon,” Marceau said. He said that the body of literature is growing, but that he does not find it to be very satisfying at this point because “there are a lot of kinks that need to be worked out.” He reiterated that he was unsure that rights were the correct way forward by giving the example that if we could do away with 90 percent of animal suffering, but did not call animals persons, we would still have won the fight to protect nonhuman animals and nature. His said that he is “open to whatever form of social change is most productive.”

Joseph Siegel: “I think there’s definitely potential for that in the future. I think it’s more accepted now than it was thirty years ago. So we’re on a trajectory; it’s going to take time. I think there’s hope that it will become more accepted.”

Steven Wise: “Oh yeah. It already is; we’re already past that. For example, in a case, we wanted to show that the argument is not frivolous and we actually cited over one hundred law review articles and legal books that had discussed our work. Some of them criticize us; some of them agree. We’re in mainstream legal scholarship. We also keep track of the discussions of our work in the press. We average, over a year, eight mentions in the press somewhere in the world every day. So every year we have 3,000, 4,000 articles that are all about us just somewhere in the world, all over the world. Some of the work is our own, but other work is coming out in Russia, Japan, and Hong Kong.”

Is there anything else you want to add?
Michael Gerard: “If the question is for photographs, in my view, it’s completely irrelevant. You know, show me pictures of cute chimpanzees or mother [and] child chimpanzees or something like that. It may be adorable, but it’s not going to change my legal analysis.”

Paul Gallay: Gallay believes that imagery will be important in increasing awareness of an idea that is not necessarily easy for a lot of people to grasp, especially because “in traditional Abrahamic religion, humankind is seen as being above nature.” He thinks that images may play an important role in helping spread awareness of the movement that is trying to “embed humankind within a broader ecology.”

Photographer
Do you think negative or positive images have a stronger impact on people’s behavior? Why?
Jo-Anne McArthur: “I think that the positive and the negative are necessary. They are absolutely necessary out in the public eye to create change and to foster thinking. I have seen people change their patterns and habits of consumption from seeing both images and also hearing the stories behind them. Sometimes images speak for themselves, and then sometimes the story behind the images is the thing that propels people into deeper thinking and deeper analysis. Different images will reach different people at different times in their lives and based on all sorts of actors in their life. It depends on their personalities as well. Some people are sort of born or raised to be able to look at suffering. I think it’s a different journey for
everybody. Sometimes you need a bit of hand-holding with negative images. I’m personally very affected by negative images and so I assume others would react the same way, but they don’t. And so, often negative images would just paralyze people. What is needed is empathy towards the person who’s viewing the image but also an action plan or some hand-holding if they want it.”

Have you ever had a specific image or series that really got people’s attention? What about those visuals do you think was particularly striking?

Jo-Anne McArthur: McArthur identified two main components to her images that have captured people’s attention. The first is that they have a close vantage point of the subject so that there is a point of connection with the animal. Typically, the animal is making eye contact with the viewer. Eye contact creates a connection with the viewer because the animal looks “into the camera and therefore out of the picture at anyone looking at them.” She also pointed out how in the wild, it is quite rare to have the opportunity to hold eye contact with an animal, but a photograph of an animal looking into the camera allows humans to sustain eye contact, which is an atypical experience. The second component is that the images were shot with a wide angle, so there is context and the viewer can place what is happening. She used the example of a photo she took of a rabbit in line for slaughter: The viewer can see a rabbit in a crate with its ears back, a line of rabbits hanging upside down, and the slaughter worker cutting the throat of the next rabbit so that all of the elements are seen and understood. McArthur described it as “beautifully rendered; it is an artistic image and even though it’s terrifying in many ways, it’s a compelling image so the viewer wants to return to the image to look at all of the components. Successful images beg the viewer to return to the subject matter.”

Building on your last response, what impact do you think looking into an animal’s eyes has on a viewer?

Jo-Anne McArthur: “It’s not something that we ever get to do very often. Certainly not, you know, out in the world. If they’re wild animals, they’re doing their own thing and they have no need really to look into our eyes. I think they have a need to perhaps look into our eyes when they need something from us, whether they are trying to say something or trying to connect, or trying to gain information. I see that a lot when I’m at factory farms, at places where animals are captive, because they’re looking for information. I guess the point is it’s a rare thing to be able to connect with them. With humans, [we] connect one-on-one with one another but so rarely with other animals. I think it probably takes us aback and it’s interesting for us when humans hold eye contact. And so, [it is] especially interesting when we hold eye contact with animals in a photograph.”

DISCUSSION
ANALYZING INTERVIEW ANSWERS

Despite having conducted interviews with only seven lawyers, the answers received were quite different with regards to what the future of these concepts is, what the purpose of advocating for creating rights for nonhuman nature is, and whether the focus should be placed on rights for ecosystems or rights for animals similar to people.

Each of the lawyers gave their perspective on whether creating rights for nonhuman animals and nature could become mainstream in legal scholarship, but there were widely differing beliefs about how accepted the ideas are now and where they will go. There were extremes in either direction; Gerrard stated that he believes that “instead of appealing to 0.1 percent of the population they will appeal to 0.2 percent of the population,” whereas Wise stated that they are already part of the mainstream and highlighted that the NhRP is mentioned in “3,000, 4,000 articles that are all about [their work]” around the world every year. These two different perspectives suggest that there is work to be done in reconciling how environmental lawyers approach the idea of creating rights for nonhuman nature.

The lawyers also differed on the purpose of advocating to create rights for nonhuman nature: Is it to be approached as a basic rights issue or as a tool to be used to raise awareness and reduce suffering? Wise described the work he does as akin to that done by human rights lawyers, while Marceau stated that if we could do away with 90 percent of animal suffering, but did not call animals persons, we would still have won the fight to protect nonhuman animals and nature. He was “open to whatever form of social change is most productive.” David Ebert echoed a similar sentiment in that he believes that arguing for rights may be a means to decrease nonhuman animal suffering and improve the quality of ecosystems even if the rights are never really granted. Phillip Karmel and Joe Siegel
both saw education as one of the more useful attributes of a discussion around granting nonhuman nature rights; both lawyers placed a great deal of weight on the importance of educating children and adults on the harm being done to the environment. The different points of view expressed bring up questions of how accepted these theories are and whether it is necessary that they succeed if the end goal could be achieved through other means like working within existing frameworks.

When asked about the similarity argument as compared to a more ecosystem-centered approach, the lawyers once again diverged. Gallay supported rights for ecosystems over rights for charismatic megafauna, while Karmel supported protecting ecosystems, but by using habitat protection plans under established procedures rather than through creating rights for ecosystems. Marceau was wary of focusing on animals similar to people, but recognized the appeal that the approach could have. Siegel was in favor of using the similarity argument insofar as it led to the education of more people on environmental issues. In other words, there was little agreement over whether the similarity argument should be used and if it were to be used, the reasons behind why it was being used over another argument.

The differing responses by this small group of environmental lawyers suggest that even among those who are working to protect the well-being of planet, there is a wide-ranging set of ideas on the best method to achieve the end goal of ensuring humans coexist with nonhuman nature.

When comparing the responses of the lawyers interviewed to the survey responses of law students, it seems as though there is a divide between the two groups: law students seem to be more supportive of the idea to give nature legal rights than the practicing lawyers who were interviewed. Across all variations of the survey, the highest percentages of answers were in the “Strongly Agree” and “Agree” response categories. This suggests that current law students are more open to the idea of creating legal rights for nature than practicing lawyers. The lawyers tended to either be open to the idea, but in the sense that they were either interested in exploring any option that could lead to protecting the environment or viewed the prospect of creating rights as unrealistic.

**THE SURVEY APPROACHES AND METHODS**

According to Fink (2013), digital surveys are best suited for populations that are spread over a wide geographic area, include a large number of people who have valid email addresses that they use regularly, and for when the study needs to include photographs or other media. Law students enrolled at American-Bar Association (ABA) approved institutions were the target population for the digital survey (and will be referred to as “law students”). Law schools are distributed across the United States in a wide geographic range. Law students are a large population: there were 203 ABA approved law schools in the United States as of 2018 and according to data from the ABA website, there were 111,561 students enrolled in law school in 2018. Law students are all assigned institutional email addresses which they are expected to regularly check to successfully complete their studies, meaning they would be able to easily access an online survey. Lastly, this study focuses on assessing photographs themselves, and it was essential that photographs could be included in the survey. These characteristics made an online survey a good medium for this study. The survey was conducted using Qualtrics, an online data collection software.

Law students were selected as the target population for several reasons, one being that the fixed population size and readily available data on law student population size from the ABA will facilitate data analysis. Secondly, law students are being educated in the modes of thought of the current United States legal and governance system, and may pursue careers in areas that could be affected by the three theories being investigated. Therefore, their background knowledge and the fact that they may eventually have influence in the field of law and the ability to institute changes to current governance systems make them a good study population. They are also more likely to be interested in ideas concerning legal systems than the average person.

150 of the 203 ABA law schools were selected to be contacted with a link to the survey. Excel was used to randomize the list of ABA law-approved law schools from the “509 Required Disclosures” page on the ABA website. The 2018 “J.D. Enrollment and Ethnicity Data” was used to get the grand total of J.D. enrollment at each school.

Emails were sent to the faculty member in a dean of students role at each of the 150 law schools and to student leaders at schools where their infor-
mation was listed publicly. The email contained an introduction to the survey and a link to the survey itself along with a request to forward the survey to their respective student bodies. The message that was sent to law schools can be found in Appendix II.

On March 3, 2019, 211 emails were sent out. From those 211 emails, thirty-six responses were received; twelve recipients agreed to distribute the survey while twenty-four declined to participate in this study. On March 10, 2019, a follow-up message was sent to the 175 people who had received the first message and not responded. From those 175 emails, twenty-six new responses were received; eleven recipients forwarded the survey to current students at their institution, while fifteen declined to participate. The survey was closed to responses on March 24; it accepted responses for three weeks. In total, there were sixty-two responses to the emails sent out, making for a response rate of 29.4 percent.

Table 1 breaks down the rate of responses to emails sent out requesting administrators and student leaders to send the survey to current students at their respective institutions. It is organized by overall response rate, positive responses, and negative responses.

<table>
<thead>
<tr>
<th>Type of Response</th>
<th>Number of Emails</th>
<th>Response Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall responses</td>
<td>62 out of 211</td>
<td>29.40 percent</td>
</tr>
<tr>
<td>Positive responses</td>
<td>23 out of 211</td>
<td>10.90 percent</td>
</tr>
<tr>
<td>Negative responses</td>
<td>39 out of 211</td>
<td>18.50 percent</td>
</tr>
</tbody>
</table>

Table 1: Survey response rates

The relatively low response rate reinforces the decision to reach out to a large pool of law schools. In order to increase response rate, there were five $20 Amazon gift cards which recipients could win from a raffle after the survey was done accepting responses. Incentives help to provide a motivation for members of the target population who might not be interested in the subject matter to complete the survey, thereby increasing response rates.

The raffle was conducted on April 5, 2019, and the message that was sent to the five winners can be seen in Appendix III. 228 respondents chose to provide their email to be entered into the raffle. Winners were chosen by randomly selecting five numbers between 1 and 228 and choosing the email that corresponded with the chosen number in the spreadsheet of the data.

EXPERIMENTAL DESIGN

The experimental design is an adapted version of the design used by Kaewnopparat (2017) which had a similar study structure, though it was on the topic of tourism and photography. The survey consisted of three parts. The first part was a description of the study that informed the participant who conducted the survey and why. It also stated that they were able to skip any questions they preferred not to answer. The third part contained demographic questions.

The second part of the survey had three different versions or blocks: the control group, experimental group A, and experimental group B. Qualtrics allows the three blocks within the second section to be randomized and presented evenly to participants so the different versions were accessed an equal number of times. Each of these three versions contained the same set of statements that aimed to capture the core ideas of both Earth jurisprudence and the NhRP, but they contained different, if any, visuals. The control group contained no visuals. Experimental group A contained photographs that matched the statements: For example, the statement “Jellyfish should have the right to exist” was paired with a photo of a jellyfish. Experimental group B contained photographs that did not match the statements: For example, the statement about a jellyfish was paired with a photograph of a landscape. The aim of the three variations was to assess whether it was the photographs themselves that influenced how participants responded to the given statements, since the same principles would be represented while the visuals varied between versions.

Whether a person processes information primarily through visual or other cues is important to acknowledge because it seems as though it could influence the results of the survey by influencing how people learn, recall information, and problem solve. If someone does not rely on images in their thinking process, photographs may have very little impact on them. However, Kaewnopparat (2017) included a section in his survey to assess if an individual processes information visually and found that “no matter whether individuals have either high or low visual styles of processing, they also received the same effects from the different color tones of photographs” and that “there seems to be no significant difference in the responses of visual style of processing with regard to color tones.” In other words, the study suggests that people
reacted similarly to the images they were presented with regardless of how visually they think. Considering these findings, such a section was omitted from the survey in this study in an effort to keep it as short as possible to increase completion rates.

SURVEY QUESTIONS
There were thirty questions in the main section of the survey and seven in the demographic section. A progress bar was included so participants had a sense of how much of the survey remained as they worked through it. The questions were primarily closed questions, meaning that a list of acceptable answers was given below the question. They were closed because the format helps provide more reliable and relevant answers and makes it easier for participants to complete the survey on any device, including a cell phone or tablet. Questions within each of the three parts appeared in a different and random order for each respondent so that question order did not influence results as done in Cameron and Williams (2015).

The questions were designed to minimize confusion for participants. Fink (2013) suggests writing questions at a sixth to eighth grade reading level to ensure that participants understand what is being asked. In an effort to meet that standard, the survey was checked using Microsoft Word’s readability statistics tool, which gives a Flesch-Kincaid score for a document’s readability. It compares the number of syllables and number of words to see what level of schooling in the United States education system a person would need to be able to easily understand a text. The survey had a 9.0 Flesch-Kincaid Grade Level score, but given that the target population is pursuing an advanced degree, it was accepted that the score was slightly higher than recommended in Fink (2013).

All of the questions in the second part asked respondents to place their responses on Likert scales, a type of ordinal scale that asks respondents to place their answer somewhere between, in this case, “Strongly Disagree” to “Strongly Agree.”

The questions being posed in the second part aimed to determine if photographs can symbolize the principles behind the adoption of a legal structure which gives nature and nonhuman animals more rights than they currently have. The style of statements is based on those used in Kaewnopparat (2017). The questions targeting Earth jurisprudence principles were based on two lists of principles from Cormac Cullinan’s 2011 edition of “Wild Law: A Manifesto for Earth Justice.” The first list, titled “Learning from Indigenous Peoples,” looks at Indigenous peoples’ cosmologies that intertwine nature and humans, while the second list summarizes the principles that Cullinan put forth over the course of the book. The questions targeting NhRP principles came from the ideas put forth in Wise (2000), Wise (2002), and the NhRP website. The survey as it appeared to respondents can be viewed in Appendix I.

A sample part two question can be seen in the figure below.

Figure 3: Questions in the second part followed the general format demonstrated above

The third part, which asked for demographic information, contained primarily closed questions on variables such as age or year of study. Given that personal questions contain sensitive information, there was a statement assuring confidentiality and the questions were based on previously tested demographic questions in other studies. Respondents’ names were not collected and emails were only collected if the participant chose to be entered into the raffle. To preserve anonymity, data collection was set up so that the main survey ended and the participant was redirected to another webpage that provided the option to enter their email. The link took the participant into a second survey so the data were stored separately in Qualtrics. To the user it flowed seamlessly because all they had to do was click, but the data were saved and stored as two separate surveys.
PHOTOGRAPH SELECTION
Since each of the questions in experimental groups A and B was shown with a photograph, there were thirty photographs in total. The photographs were paired with a question in their version and always appeared with that same question regardless of where that question appeared in that version of the survey. The photographs show landscapes, landscapes with humans and/or animals in them, and portraits of animals. Most of the photographs are in color because Labao et al. (2007) found that color photographs increase the value attached to an environmental good depicted in the photograph more than black and white photographs.

The photographs in the survey were created by professional nature photographers who aim to play a role in conservation efforts and foster a connection between the viewer and the natural world. Photographers include Jimmy Chin, Keith Ladzinski, Frans Lanting, Peter McBride, Jo-Anne McArthur, Cristina Mittermeier, Thomas Peschak, Joel Sartore, Sean Scott, Brian Skerry, Brent Stirton, and Carleton Watkins. All images were used with permission from the copyright holders. Each of these photographers is highly accomplished, having received international recognition over the course of their respective careers.

The overarching criteria for all of the photographs in the survey were based on the results of the studies below that aim to determine what aspects of nature inspire positive reactions and which situations inspire indignation in people. These two emotional reactions were chosen because emotional connection to nature and moral outrage against harm to nature are powerful motivators for pro-environmental behavior. Two sets of specific criteria were developed to select the different kinds of photographs in the survey; one set of criteria for the landscape photographs and another for the photographs of animals. Both are explained in depth below.

LANDSCAPE PHOTOGRAPH CRITERIA
The landscape photographs included some images showing only plants and inanimate nature. Carleton Watkins’s legacy of an understanding of wilderness as separate from humans is problematic in the Earth jurisprudence framework which tries to foster a connection between people and nature. In two critical analyses by Deluca and Demo (2000) and (2001), the authors argue that Watkins created a power dynamic between humans and nature and contributed to the erasure of the people who lived in Yosemite in the narrative he created. Many of his photographs like the image in Figure 4 place the viewer in a position of power in relation to, for instance, a valley that is lent a sense of endlessness that is simultaneously conquered by the person looking over the vast space. Watkins’s photographs also create a dichotomy between humans and nature by showing only what appears to be pristine nature without any human presence. Watkins framed his photographs to show untouched nature and made the decision not to photograph any indicators that the valley had been inhabited by Native Americans who were forced out of the area or that there were workers in the park every day. Nevertheless, his work was essential to the establishment of how American landscapes are seen today and in creating a “representational vocabulary” for preservationist environmental movements where images of vast wild spaces are associated with pristine beauty and escape. Considering his large contribution to the field of landscape photography and the use of landscape photos in the survey, a small number of images created by him or in his style were chosen to be included in the survey.

Figure 4: “Yosemite Valley from the Best General View” by Carleton Watkins, 1866

The pool of landscape photographs contained a mixture of pristine locations and locations that have experienced human-caused harm because Earth jurisprudence is centered on the idea that humans are part of nature and need to embrace that connection for the well-being of all parties. Comparing the responses to the photographs intended to evoke positive emotions and those intended to evoke feelings of shame and guilt may help better understand what type of photo inspires greater support for expanding the definition of legal personhood.
ANIMAL PHOTOGRAPH CRITERIA

According to Yale professor Stephen Kellert, four central elements affect how people view animals: prior outlook on the species, past familiarity with the species, the role of the species in human life, and human perceptions of the species’ ability to learn and perceptions of its aesthetic value. Humans respond positively to animals that have a “relatively large head, flat face and large eyes.” Czech et al. (1998) found that people often like plants, birds, mammals, and fish more than other animals. It is also possible to predict which animals people will react to negatively. Knight (2008) found that cultural associations with animals may impact how people view them; for example, people fear bats because of their association with vampires. Another study found that humans have negative responses to organisms substantially different from themselves, like invertebrates, and that there is a connection between human preference for an animal and its similarity to humans.

Based on the findings of the studies above, nonhuman animal photographs included used primarily organisms that most U.S.-based students should be familiar with, should have positive cultural associations with, that are colorful and pleasant to view, that have known ecological importance, and that share some characteristics with people e.g. birds, big cats, whales, and elephants. To match the spirit of Earth jurisprudence and foster appreciation for all aspects of nature, photographs of nonhuman animals that did not meet the criteria were included as well, though there were a smaller number of those images.

Another important aspect of nonhuman animal photography is the role that nonhuman animals’ eyes play in how they impact people. In his essay “Why Look at Animals?” art critic John Berger writes that part of what makes interacting with nonhuman animals unique is that they are capable of looking back at people. They appear to register a perception of the human they see, giving them a certain power “comparable with human power, but never coinciding with it.” This power is illustrated by Aldo Leopold’s description of his experience killing a wolf; he got to “the old wolf in time to watch a fierce green fire dying in her eyes,” realized that she knew more than he ever would about the mountain where he killed her, and that it had been a mistake to shoot such an animal for the sake of the pleasure he gained from hunting. The sway held by looking into a nonhuman animal’s eyes adds a layer of depth to how people view nonhuman animals and consequently, several photographs were included where the nonhuman animal is looking straight at the viewer.

In an attempt to determine whether participants react more strongly to photographs that evoke positive emotions or emotions of shame and guilt, some of the animal photographs were healthy members of the species that the above studies identified as well-liked, and some were unhealthy, imperiled, or dead members of those species. For example, there was a healthy orca included, as well as a pod of orcas with a burning oil rig in the background.

Table 2 below contains a breakdown of how many photographs in the pool of thirty were meant to evoke which kind of emotions in the viewer. The proportion of photographs was skewed toward those that foster a connection with nature because connecting with nature has been shown to reinforce pro-environmental behavior more consistently. More specifically, Wilson (1984) put forth the theory of biophilia which says that all humans have an innate tendency to connect with nature and other organisms. Similarly, Kals and Mueller (2012) found that affective connection to nature, defined as a positive emotional bond with nature, leads to people experiencing positive emotions when they interact with nature and engaging in more pro-environmental behavior. As a result, it seemed likely that viewing photographs meant to evoke positive feelings toward nature would tap into these emotions and lead to greater support for creating legal rights for nature.

Negative emotions have also been shown to motivate pro-environmental behavior but they are less predictable; the range of negative emotions like, fear, threat, indignation, and sadness can evoke different responses. For instance, “research on models of pro-environmental action that incorporated feelings of fear and threat [show] consistently that these emotions do not play an important role in explaining pro-environmental commitment and behavior.” Likewise, there is an inverted association between anxiety and pro-environmental behavior; too much anxiety and people can go into denial to avoid panic. However, invoking sadness has been found to increase pro-environmental behavior. Overall, there are “differences in the relative power of the emotions as was shown by the various forms of indignation,” and the greater possibility of inciting a reaction that would decrease pro-environmental behavior led to the inclusion of fewer photos meant to evoke negative emotions.
PILOT TESTING
In collaboration with graduate students Xiaoyun Qin and Zhenhuan Xie of the Columbia University Statistics Department consulting service, R and Excel, software programs, were used to analyze the data. The first step taken was to analyze the demographic information collected to better understand the population which took the survey. An ANOVA test and a Tukey test were used to understand how responses changed when the experimental variables changed.

The categorical variables were coded into numbers so that scores could be calculated for the responses. Strongly disagree was coded as one and strongly agree was coded as five. The data were separated into three categories so that R could better process it: the three versions (experimental group A, experimental group B, and the control group) were respectively coded as A, B, and C. The three categories of pictures (intended to evoke positive responses, intended to evoke negative responses, and no pictures) were coded as P, N, and O, respectively. Earth jurisprudence and NhRP approach were coded as E and F. A 0.05 significance level was used to assess whether results were statistically significant.

After the data were cleaned, R was used to conduct a two-way ANOVA test. This test looks at how two categorical independent variables influence a continuous dependent variable, in this case, how the survey version and type of picture influence a participant’s responses to the questions asked. It also considers if there is any interaction between the independent variables. In this case that is the version and the type of picture. Interaction is when one variable influences the other.

In order to assess whether the survey sample is representative of the overall population of law students in the US, analysis was conducted based on available demographic data from the ABA and law schools on their student populations.

![Figure 6: The figure above shows the breakdown of respondents by their self-reported school year. The data on school year were calculated using the responses of 230 participants who answered the question posed. There was a slightly higher percentage of first years who responded to the survey.](image)

RESULTS
SURVEY DATA ANALYSIS
The survey was open for twenty-one days from March 3, 2019. It got a total of 289 responses. Not all of the 289 respondents completed all thirty of the part two questions and after removing responses that had not answered all thirty of the main questions, 240 were left for analysis.

Not all respondents who completed the thirty questions in the main section answered all of the demographic questions. The average age of respondents was 27.9 years, the median age was twenty-six years, the mode was twenty-four years, the lowest age was twenty, and the highest age was sixty-four. The age breakdown was calculated using data from 234 responses to the question on age. Below is Figure 6 which shows the school year of respondents.

![School Year of Respondents (%)](image)

Table 4 below shows the responses of the 226 people who answered the question “Where do you attend law school?” There were responses from twenty-three law schools across the United States. The number of responses from each school varied greatly. The University of Nevada- Las Vegas had the highest number of respondents, with fifty-one students completing the survey, while the lowest number of respondents was tied at five universities that each only had one respondent complete the survey. The table also contains the numbers of pins that correspond to where each law school is in the US as shown in Figure 7.
Table 4: Responses by law school attended

<table>
<thead>
<tr>
<th>Rank number on map</th>
<th>Law school attended</th>
<th>Number of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>American University</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>Charleston School of Law</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>Columbia University</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>Emory University</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>Fordham University</td>
<td>26</td>
</tr>
<tr>
<td>6</td>
<td>George Mason University</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>Gonzaga University</td>
<td>15</td>
</tr>
<tr>
<td>8</td>
<td>Lewis and Clark College</td>
<td>17</td>
</tr>
<tr>
<td>9</td>
<td>Mitchell Hamline</td>
<td>23</td>
</tr>
<tr>
<td>10</td>
<td>New York University</td>
<td>12</td>
</tr>
<tr>
<td>11</td>
<td>Ohio State University</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>Saint Louis University</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>Santa Clara University</td>
<td>3</td>
</tr>
<tr>
<td>14</td>
<td>Texas A&amp;M University</td>
<td>13</td>
</tr>
<tr>
<td>15</td>
<td>University of Chicago</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>University of Georgia</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>University of Kansas</td>
<td>6</td>
</tr>
<tr>
<td>18</td>
<td>University of Minnesota</td>
<td>4</td>
</tr>
<tr>
<td>19</td>
<td>University of Nevada-Las Vegas</td>
<td>51</td>
</tr>
<tr>
<td>20</td>
<td>University of San Diego</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>Valparaiso University</td>
<td>13</td>
</tr>
<tr>
<td>22</td>
<td>West Virginia University</td>
<td>12</td>
</tr>
<tr>
<td>23</td>
<td>Yale University</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 4: Responses by law school attended

Figure 7: The map above shows where each law school is geographically located. Law schools were spread across the US, though there were more schools in the eastern part of the country that participated than in the west.

Figure 8 shows the term with which respondents chose to identify their gender.

Gender Identity of Respondents (%)

![Gender Identity Chart]

Figure 8: The graph above shows the responses of the 237 people who answered the question “Which term do you use to describe your gender identity?” There were more women than men who completed the survey and a small percentage of respondents who chose to answer non-binary, prefer not to say, or other.

The ABA provides the gender breakdown of first year classes, but does not provide up-to-date gender information after the first year. In order to get an estimate of the current gender breakdown at law schools, data was downloaded from the ABA website on 1L class composition from 2016-2018; it can be viewed in Table 5. The first year classes of 2016-2018 are the current law students in the country so even though there may have been some attrition since the statistics were reported, it is an accurate estimate of the proportion of men and women currently attending law school at an ABA approved institution.

<table>
<thead>
<tr>
<th>First Year (FY) Class</th>
<th>Total FY Class</th>
<th>Total FY Class</th>
<th>Total FY Class</th>
<th>Total FY Class</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Men</td>
<td>Women</td>
<td>Other</td>
<td>Other</td>
</tr>
<tr>
<td>2018</td>
<td>17,977</td>
<td>20,366</td>
<td>47</td>
<td>38,390</td>
</tr>
<tr>
<td>2017</td>
<td>19,589</td>
<td>21,523</td>
<td>24</td>
<td>41,136</td>
</tr>
<tr>
<td>2016</td>
<td>18057</td>
<td>19032</td>
<td>17</td>
<td>37106</td>
</tr>
<tr>
<td>SUM</td>
<td>55,623</td>
<td>60,921</td>
<td>88</td>
<td>116,332</td>
</tr>
</tbody>
</table>

% of Law Students

| % of Law Students | 47.69 | 52.23 | 0.08 | 100% |

Table 5: ABA gender data for law school classes

Respondents were asked to self-identity as color-blind; 236 respondents chose to answer the question “Are you color blind?” Of those responses, 1.30 percent said they were color-blind and 98.70 percent said they do not identify as color blind. Respondents
were also asked to provide their favorite color from a list of seven options. However, the information was not used during analysis and so, was omitted.

Respondents were asked to answer the question “Do you intend to pursue a career related to environmental law? If no, please list the area of law you intend to pursue.” Of the 237 participants who answered the question, 16.50 percent did intend to pursue environmental law, 60.30 percent did not intend to pursue environmental law, and 23.20 percent were undecided. Table 6 shows the responses students filled in for the area of law they were interested in pursuing if they did not want to pursue environmental law; 103 respondents listed legal various subject matter areas they intend to pursue.

<table>
<thead>
<tr>
<th>Area of Law</th>
<th>Number of Respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture law</td>
<td>1</td>
</tr>
<tr>
<td>Animal Law</td>
<td>3</td>
</tr>
<tr>
<td>Bankruptcy</td>
<td>1</td>
</tr>
<tr>
<td>Business Law</td>
<td>2</td>
</tr>
<tr>
<td>Cannabis Law (administrative)</td>
<td>1</td>
</tr>
<tr>
<td>Capital markets</td>
<td>1</td>
</tr>
<tr>
<td>Child protection law</td>
<td>4</td>
</tr>
<tr>
<td>Church law</td>
<td>1</td>
</tr>
<tr>
<td>Civil law</td>
<td>3</td>
</tr>
<tr>
<td>Class action litigation</td>
<td>1</td>
</tr>
<tr>
<td>Commercial litigation</td>
<td>5</td>
</tr>
<tr>
<td>Construction law</td>
<td>1</td>
</tr>
<tr>
<td>Corporate law</td>
<td>8</td>
</tr>
<tr>
<td>General criminal law</td>
<td>12</td>
</tr>
<tr>
<td>Criminal defense law</td>
<td>4</td>
</tr>
<tr>
<td>Criminal prosecution</td>
<td>4</td>
</tr>
<tr>
<td>Employment law</td>
<td>3</td>
</tr>
<tr>
<td>Energy and natural resources law</td>
<td>2</td>
</tr>
<tr>
<td>Entertainment law</td>
<td>1</td>
</tr>
<tr>
<td>Family law</td>
<td>5</td>
</tr>
<tr>
<td>Gaming law</td>
<td>3</td>
</tr>
<tr>
<td>Healthcare law</td>
<td>3</td>
</tr>
<tr>
<td>Human rights</td>
<td>3</td>
</tr>
<tr>
<td>Immigration law</td>
<td>6</td>
</tr>
<tr>
<td>Intellectual property law</td>
<td>4</td>
</tr>
<tr>
<td>International environmental law</td>
<td>1</td>
</tr>
<tr>
<td>International law</td>
<td>2</td>
</tr>
<tr>
<td>LGBT rights</td>
<td>1</td>
</tr>
<tr>
<td>Private equity</td>
<td>1</td>
</tr>
<tr>
<td>Personal injury</td>
<td>3</td>
</tr>
<tr>
<td>Plaintiff side litigation</td>
<td>1</td>
</tr>
<tr>
<td>Public interest</td>
<td>4</td>
</tr>
<tr>
<td>Public international law</td>
<td>1</td>
</tr>
<tr>
<td>Real estate</td>
<td>2</td>
</tr>
<tr>
<td>Riparian rights</td>
<td>1</td>
</tr>
<tr>
<td>Tax law</td>
<td>2</td>
</tr>
<tr>
<td>Tenant representation in landlord-tenant suits</td>
<td>1</td>
</tr>
<tr>
<td>Transactional nature</td>
<td>1</td>
</tr>
<tr>
<td>Undecided, but not environmental</td>
<td>2</td>
</tr>
</tbody>
</table>

There were ten types of questions in the survey due to the three variations on photographs, two different theories, and three versions. For example, one question type was positive photographs shown with Earth jurisprudence questions that match the statements. However, there were few differences between responses to Versions A and B (when the photographs did or did not match the statements) and so the data from those two versions was combined in the figures and analysis. Figures 9 and 10 below show the breakdown of respondents’ answers to each type of question by percent.

Version C of the survey contained eighteen questions related to Earth jurisprudence. It was accessed by eighty participants. Versions A and B contained nine questions about Earth jurisprudence shown with negative photographs and twenty-seven questions shown with positive photographs when they were combined. 160 respondents accessed these questions. Figure 9 shows an overall breakdown of participants’ responses to the different Earth jurisprudence question types.

![Figure 9](image)

**Figure 9:** The graph above shows the responses of participants to Earth jurisprudence questions. For Version A+B (positive photos), “Agree” received the most responses. For Version A+B (negative photos), “Strongly agree” received the most responses. For Version C (no photos), “Agree” received the most responses.
Version C contained twelve questions about the NhRP that were shown with no photographs and 80 respondents answered these questions. Versions A and B contained nine questions about the NhRP shown with negative photographs and 15 questions shown with positive photographs when they were combined. 160 respondents accessed these questions. Figure 10 shows an overall breakdown of participants’ responses to the different NhRP question types.

Figure 10: The graph above shows the responses of participants to the NhRP questions. For Version A+B (positive photos), “Agree” received the most responses. For Version A+B (negative photos), “Agree” received the most responses. For Version C (no photos), “Agree” received the most responses.

The two-way ANOVA test conducted on the survey results showed that there was a statistically significant relationship between the version, type of picture, and a person’s response to the questions for both the questions testing Earth jurisprudence ideas and the questions testing the Nonhuman Rights Project approach. For Earth jurisprudence questions, there was a statistically significant interaction between the version and type of picture while for the Nonhuman Rights Project questions the interaction was not statistically significant.

If the two-way ANOVA yielded significant relationships, a Tukey test was conducted to test the differences between means within a given category by comparing all the possible pairs of the means. The Tukey test involved comparing the mean scores of the responses to questions with no pictures, questions with pictures intended to evoke positive emotions, and questions intended to evoke negative emotions.

Table 7 shows the mean differences obtained by performing a Tukey test on the different types of photographs paired with Earth jurisprudence questions. The different options were (No photos, Version C), (Negative photos, Version A+B), and (Positive photos, Version A+B).

<table>
<thead>
<tr>
<th>Mean difference</th>
<th>Lower bound</th>
<th>Upper bound</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(No photos, Version C) - (Negative photos, Version A+B)</td>
<td>-0.18</td>
<td>-0.31</td>
<td>-0.04</td>
</tr>
<tr>
<td>(Positive photos, Version A+B) - (Negative photos, Version A+B)</td>
<td>-0.23</td>
<td>-0.36</td>
<td>-0.11</td>
</tr>
<tr>
<td>(Positive photos, Version A+B) - (No photos, Version C)</td>
<td>-0.06</td>
<td>-0.16</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Table 7: Tukey test results

The Tukey test showed that for Earth jurisprudence questions the mean score of responses to questions shown with photographs intended to evoke negative emotions was higher than the mean score of responses to questions shown with no photographs (No photos, Version C - Negative photos, Version A+B). In other words, respondents had more positive responses towards Earth jurisprudence ideas when presented with negative photographs than when no photographs were present. Additionally, respondents had more positive responses when presented with photographs meant to evoke negative emotions than when presented with photographs meant to evoke positive emotions (Positive photos, Version A+B - Negative photos, Version A+B). The mean difference for the (Positive photos, Version A+B - No photos, Version C) category was not statistically significant. Figure 11 below represents the results of the Tukey test from Table 7 graphically.
Figure 11: The plot depicts the mean difference between the three types of photographs that participants saw with Earth jurisprudence questions. The data show that photos meant to evoke negative responses got the most positive responses followed by no photos and with photos meant to evoke positive responses having the least positive responses. In short, photographs meant to evoke negative emotions received the most positive responses from participants.

Table 8, below, shows the mean differences obtained by performing a Tukey test on the different types of photographs that appeared with the NhRP questions.

<table>
<thead>
<tr>
<th>Category</th>
<th>Mean difference</th>
<th>Lower bound</th>
<th>Upper bound</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(No photos, Version C) - (Negative photos, Version A+B)</td>
<td>0.11</td>
<td>-0.04</td>
<td>0.26</td>
<td>0.19</td>
</tr>
<tr>
<td>(Positive photos, Version A+B) - (Negative photos, Version A+B)</td>
<td>0.17</td>
<td>0.04</td>
<td>0.31</td>
<td>0.01</td>
</tr>
<tr>
<td>(Positive photos, Version A+B) - (No photos, Version C)</td>
<td>0.07</td>
<td>-0.06</td>
<td>0.20</td>
<td>0.46</td>
</tr>
</tbody>
</table>

Table 8: Tukey test conducted on different types of photographs with the NhRP questions

The Tukey test showed that for NhRP questions, the (No photos, Version C - Negative photos, Version A+B) category had no statistical significance. The mean score of responses to questions shown with photographs meant to evoke positive emotions was higher than the mean score of responses shown with photographs intended to evoke negative emotions (Positive photos, Version A+B - Negative photos, Version A+B). This means that respondents had more positive reactions when presented with photographs meant to evoke negative emotions than when presented with photographs meant to evoke positive emotions. The mean difference of the (Positive photos, Version A+B) - (No photos, Version C) category was not significant. Figure 12 below represents the results of the Tukey test from Table 8 graphically.

Figure 12: The plot depicts the mean difference between the three types of photographs that participants saw with NhRP questions. The data show that photos meant to evoke negative responses had the most positive responses. The two other categories did not have statistically significant results. In summary, photographs meant to evoke negative emotions received the most positive responses from participants.

A Tukey test was also done to compare the differences between the mean scores within the version category. Below, Table 9 shows the results of a Tukey test conducted on the different versions of the survey for Earth jurisprudence questions.

Table 9: Tukey test conducted on the different versions of the survey for Earth jurisprudence questions

<table>
<thead>
<tr>
<th>Category</th>
<th>Mean difference</th>
<th>Lower bound</th>
<th>Upper bound</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version B - Version A</td>
<td>-0.01</td>
<td>-0.12</td>
<td>0.10</td>
<td>0.99</td>
</tr>
<tr>
<td>Version C - Version A</td>
<td>0.20</td>
<td>0.09</td>
<td>0.31</td>
<td>0.00010</td>
</tr>
<tr>
<td>Version C - Version B</td>
<td>0.20</td>
<td>0.09</td>
<td>0.32</td>
<td>0.00007</td>
</tr>
</tbody>
</table>

The mean difference of responses to questions in the (Version B - Version A) category was not statistically significant. The mean score of responses to questions in Version C was higher than the mean score of responses to questions in Version A (Version C - Version A) and the p-value was below .05 so it was statistically significant. This means that respondents had more positive responses when presented with no photographs than when presented with photographs that matched the statements. The mean score of responses to questions in Version C was higher than the mean score of responses to questions in Version B (Version C - Version B) and the p-value was below .05 so it was statistically significant. This means that respondents had more positive responses when presented with no photographs than when presented with photographs that do not match the statements. Figure 13 below
represents the results of the Tukey test shown in Table 9 for the version of the survey for Earth jurisprudence questions graphically.

**Figure 13:** The plot depicts the mean difference between responses to different versions of the survey for Earth jurisprudence questions. The data show that the version of the survey with no photographs received the most positive responses.

Table 10 shows the results of the Tukey test for the different versions of the survey for the Nonhuman Rights Project questions.

<table>
<thead>
<tr>
<th>Mean difference</th>
<th>Lower bound</th>
<th>Upper bound</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version B- Version A</td>
<td>-0.01</td>
<td>-0.15</td>
<td>0.12</td>
</tr>
<tr>
<td>Version C- Version A</td>
<td>0.12</td>
<td>-0.01</td>
<td>0.26</td>
</tr>
<tr>
<td>Version C- Version B</td>
<td>0.13</td>
<td>0.00</td>
<td>0.27</td>
</tr>
</tbody>
</table>

**Table 10: Tukey test conducted on the different versions of the survey for NhRP questions**

None of the p-values were below the cutoff value of 0.05. However, if a 0.10 significance level were to be used, which is still considered to be robust, the results of (Version C- Version A) and (Version C- Version B) would be significant. For (Version C- Version A) this means that responses were more positive when no photographs were presented than when photographs that matched the statements were shown. For (Version C- Version B) this means that responses were more positive when respondents were presented with no photographs than when they were presented with photographs that did not match the statements. Figure 14 below represents the results of the Tukey test shown in Table 10 for the version of the survey for NhRP questions graphically.

**Figure 14:** The plot depicts the mean difference between responses to different versions of the survey for NhRP questions. The data show that the version of the survey with no photographs received the most positive responses.

**DISCUSSION**

**UNDERSTANDING THE SURVEY’S DEMOGRAPHIC INFORMATION**

This subsection aims to understand if the sample population was representative of the greater population of law students studying at ABA approved law schools.

The population in the survey sample was composed of 38.4 percent men, 58.2 percent women, and 3.4 percent other (when combining the “non-binary”, “prefer not to say”, and “other” responses from the survey data). The broader population of law students is 48 percent men, 52 percent women, and .08 percent other. The proportion of women to men in the survey sample was slightly higher than in the greater population of law students in the US. The similarity between the gender proportions means that the sample data with regard to gender is reasonably similar to the population.

The ABA does not require law schools to disclose average age of students. In order to determine an approximate average age of law students, the first year profiles of Columbia Law and Yale Law were consulted because both provide an average age of their first year students: at Yale the average first year starting school in 2018 was twenty-five while at Columbia 61 percent of first years were between the ages of 21 and 24. This suggests that the average age of law students is around twenty-six or twenty-seven years old. The average age of the sample was 27.9, suggesting that the sample is reasonably similar to the greater population of law students.

In terms of intended area of focus, the ABA does not provide statistics on fields that graduates enter. There were, however, thirty-nine different areas of
law listed by respondents as areas they intend to pursue other than environmental law which suggests that the sample population was quite diverse in its interests and may be reflective of the greater population of law students. Only 16.5 percent of respondents answered that they intended to pursue environmental law. As there are no statistics on how many environmental lawyers there are it is impossible to know if this percentage is higher or lower than in the population, but it is low enough that it has likely not unduly influenced the results by skewing them in favor of the concepts being examined by having a population which is biased in favor of granting nature and nonhuman animals legal rights. Such a bias was a concern because the survey was partially distributed through student groups at law schools dedicated to environmental or animal law.

The question on color blindness was disregarded in the analysis because only 1.3 percent of respondents self-identified as color blind so color blindness likely did not influence the results.

Overall, it seems as though the sample has a reasonably similar composition to the population of law students, though the exact data were not publicly available and the estimates may be off.

EFFECT OF PHOTOGRAPHS ON RESPONDENTS’ ANSWERS
For both Earth jurisprudence and the NhRP approach, negative photographs evoked more positive responses than both photographs meant to evoke positive emotions and no photographs did. However, the questions with no photographs had similar results to the questions with negative photographs and the general trend seemed to be that positive photographs had a negative effect on respondents’ answers. One possible interpretation of this trend is seeing negative photographs highlighted the wrongs occurring in the environment while the positive photographs may have glazed over poor treatment of the environment and projected an image of a healthy environment that does not need extra protections, leading participants to respond more strongly to statements for protecting nature when they saw photographs meant to evoke negative emotions than when they saw photographs meant to evoke positive emotions. Such an effect could be similar to the one observed in Courchamp et al (2018); a study on ten species of animals considered very charismatic which found that there exists a biased perception of their well-being due to widespread use of their images that undercuts the immediacy of their threatened status. The interview conducted with Jo-Anne McArthur reinforces the conclusion that negative photographs can be powerful motivators for pro-environmental behavior as did the interview with David Ebert in which he explained how alleviating the suffering of animals is one of his main motivators in his work at the Animal Defense Partnership.

A second interpretation could be that with regard to questions with negative photographs and no photographs, people considered the human impact on the environment and in the context of the human impact supported creating rights for nature. Whereas when they were presented with just an image of an animal in its habitat the strangeness of the idea of creating rights for nature may have been accentuated. In other words, respondents could be in favor of rights for nature when the idea that the rights are necessary because of human actions is emphasized and they may be less supportive when the appeal is made from the point of view that animals on their own have intrinsic value. Leiserowitz et al (2018) found in a nationally representative survey that 62 percent of Americans believe that climate change is a result of anthropogenic action. The underlying tendency toward believing that humans are harming the planet could explain why the control group had more positive responses than the questions shown with positive photographs.

For both the questions looking at Earth jurisprudence ideas and the questions looking at the NhRP approach, the control group with no photographs received more positive responses than either of the two versions of the survey with photographs. It is possible that the target population of law students is a particular subset that responds poorly to imagery. In his interview, Michael Gerrard expressed his belief in the idea that lawyers and law students are more convinced by words and logic than by other means of argumentation. Gerrard said that he believes the question of photographs to be irrelevant to his legal analysis. The lack of impact of photographs on law students suggests that attempts to convince lawyers to support the creation of rights for nonhuman animals and nature could be more effective if they are primarily text-based since lawyers are trained to think through issues based on the evidence presented to them. It is possible that using text-based appeals founded in logic and evidence, rather an emotional appeal through a photograph, could be more effective in evoking positive responses from lawyers.

In contrast to the idea that text-based appeals
are most effective was Paul Gallay’s belief that photographs could help spread awareness of these theories among a more diverse population. Joseph Siegel and Philip Karmel spoke about the importance of environmental education in increasing public support for protecting the environment. Kellert (1996) showed that higher levels of education are related to higher levels of value being placed on the intrinsic value of the environment. Environmental education leads to greater familiarity with a given subject and that increased familiarity leads to people placing a greater value on a given species or landscape. It is possible that photographs could be more effective at increasing support for creating legal rights for nature and nonhuman animals among a broader population of people with graduate degrees. This population may be more open to being swayed by an emotional response and less focused than law students on words and concrete evidence.

CONCLUSIONS
SUMMARY OF FINDINGS
The results of this study shed light on the viability of using images to represent the philosophies of Earth jurisprudence, the NhRP, and the similarity argument in legal spheres. It seems as though photographs intended to evoke negative emotional reactions resonate more with law students than photographs intended to evoke positive emotional reactions, but having no photographs evokes more positive responses than having photographs. In other words, this study suggests that if photography is used as a tool to evoke more positive reactions towards these theories in law students, the most effective type of photographs are those intended to evoke emotions of moral outrage due to harm being done to the environment. Text-based appeals, however, could be the most effective method of communicating with law students.

IMPLICATIONS
There are many organizations founded to promote legal rights for nature and nonhuman animals that could use the information from this study’s survey to improve their communications strategies and to increase awareness of Earth jurisprudence and the NhRP. Some such organizations include the Center for Earth Jurisprudence, the Global Alliance for the Rights of Nature, the Earth Law Center, the Australian Earth Laws Alliance, the Gaia Foundation, the Animal Defense Partnership, We Animals, and the Nonhuman Rights Project.

Although the results could inform the marketing efforts of such organizations, the data are a step removed from marketing itself; the information could also inform an art exhibit or publication interested in these ideas. This study is a step before marketing because it does not involve the persuasion and manipulation essential to marketing. This study aimed to understand what response photographs evoke in viewers on a deeper level—it may be difficult for a person to put into words why a certain image makes them more sympathetic to Earth jurisprudence, whereas in an ad campaign, the visuals and/or text being presented have been carefully selected and manicured to guide a viewer to a given conclusion that can easily be traced.

AREAS FOR FURTHER RESEARCH
There was a wealth of data collected through the survey, however, not all of them ended up being relevant to the question the study aimed to answer and as such was not used in analysis. For instance, the question asking respondents for their favorite color was not used in analysis. As a result, if the scope of the study were to be expanded or the data revisited, it would be possible to obtain more information from the data set. Similarly, it would have been possible to create models to see how specific groups responded to the questions. For example, analysis could have been done comparing differences in responses to questions with landscape and animal photographs. It could also have been illuminating to see if a higher proportion of the respondents to the control group happened to be interested in pursuing environmental law. If it so happened that more of those interested in environmental law opened the control version, it could explain why the control group had more positive responses than the two experimental groups.

The finding that the control group received more positive responses than either version with photographs was somewhat unexpected. Further research could be conducted into whether these results were due to the specific population of law students or the design of the study. A follow-up study could be conducted where two populations, one of law students and one of a more general population, take a similarly designed survey and the responses across the different populations could be compared to see whether legal training is what made the photographs have a lesser impact than the text. A second follow-up study that could be conducted is whether the focus of respon-
dents was an issue in the design of this study; it is possible that when presented with an image and a statement rather than just a statement, respondents’ focus on the question being asked decreased and they read the statements less thoroughly, resulting in more neutral responses. Such a study could remove the competition between the photographs and the statements by presenting respondents with either just a statement or just a photograph, then asking if the respondent agrees with pursuing further environmental action on that issue.

It could also be telling to show photographs with captions that reinforce a specific statement and then show only the statements. Such a study would aim to answer the question of whether photographs are effective as illustrations combined with text for populations that favor more text-based argumentation.

To address the question of whether seeing or having an existing belief that human harm to the environment is what led to more positive responses from negative photograph and no photograph questions, a study could be conducted where negative photographs that show human harm to the environment are presented to one group of respondents, and photographs that show animals demonstrating their intelligence are shown to another group. Such a study could help resolve the uncertainty regarding what caused positive photographs to have a negative effect on respondents’ answers.

Another potential study would be to look into what motivated people who supported measures that enact Earth jurisprudence ideas to vote for or enact that measure. To date, there have been instances around the world and at least two cases in the United States where Earth jurisprudence has been enacted in different forms. For example, the preamble to Ecuador’s 2008 constitution states that the sovereign people of Ecuador commit themselves to “a new form of public coexistence, in diversity and in harmony with nature.” This is particularly important because the preamble frames the intent of the rest of the document. In 2017, New Zealand gave the ninety-mile Whanganui River the same legal status as a living person. The river has two guardians who can represent it in legal cases that concern its interests. A community in Florida passed an “ordinance with legally enforceable rights for natural communities and striped corporations that unlawfully deposited sewage sludge on farmlands of their” rights as juristic persons. The area that voted for the Lake Erie Bill of Rights, which creates the rights to exist, flourish, and naturally evolve for the body of water, could also be a site for study.

Surveying or interviewing people from those places who were in favor of those measures may also provide insight into what motivates support for ideas which seek to create legal rights for nonhuman animals and nature.

APPENDICES
APPENDIX A
Survey as it appeared to the control group
Below is the survey as it appeared to respondents in the control group. The control group’s survey was chosen because the researcher did not have permission to distribute the photographs used in the survey beyond use in the survey itself. Respondents took one of three versions of the survey. Two other versions are not shown here. The version that appeared to experimental group A included the same set of statements with photographs that matched the statements. The version that appeared to experimental group B included the same set of statements with photographs that did not match the statements. For example, the photo of a tree was paired with the statement about penguins.

This study aims to understand law students’ opinions concerning the idea that nature should have legal rights. My name is Elza Bouhassira and I am senior in the Sustainable Development program at Columbia University. I am conducting the survey as an essential part of my undergraduate senior thesis and as such greatly appreciate your willingness to participate.

Participants are asked to complete the survey that follows. Participation is voluntary and please be aware that you can skip any questions you would prefer not to answer. The survey takes on average 4-7 minutes and consists of 3 parts. All demographic information collected will be kept confidential to the best of the researcher’s ability.

At the end of the survey, participants will have the option to enter their e-mails to be put into a raffle for one of five $20 Amazon Gift Cards. E-mails will be stored separately from the rest of a participant’s responses and will only be used for the purpose of notifying winners of the raffle. Should any questions arise about the survey please reach out to the researcher at [email...
Thank you again for your willingness to participate in my project!

Clicking below indicates that I have read the description of the study and I agree to participate in the study.

I have read the introduction and agree to participate in the survey

I feel part of a larger community of animals, plants, and natural landscapes

Jellyfish should have the legal right to exist

Chimpanzees should have the legal right to exist

Cheetahs should have the legal right to live in forests

A mouse in the walls of a building has the right to life

A family of penguins in the Antarctic has the right to life

It should be illegal to carve names into trees because it is disrespectful to the trees

More natural areas should be protected from human disruptions

In an area with limited amounts of water a set amount of water should be dedicated to ensure wildlife has access to it

An area with polluted air should be cleaned up even if no humans live there

Rhinos should have the legal right to be free from torture or cruel treatment

Coral reefs that have been destroyed by humans should be legally mandated to be restored to their state prior to disruption

I would choose to eat less meat for environmental reasons (reduction of carbon and methane emissions)

At their most fundamental level laws should consider the well-being of the Earth

Humans need the diversity of ecosystems on Earth to thrive

An amendment should be made to the US constitution that recognizes the rights of the Earth and the organisms that live on it to exist and to have a habitat

I would represent an octopus whose legal right to be in a clean ocean habitat had been violated

I would represent a river whose legal right to be free from pollution had been violated

Elephants should have the legal right to bodily integri-
Great apes should have the legal right to bodily liberty (the right not to be imprisoned)
Strongly disagree  Disagree  Neither agree nor disagree  Agree  Strongly agree

Dolphins should have the legal right to bodily liberty (the right not to be imprisoned)
Strongly disagree  Disagree  Neither agree nor disagree  Agree  Strongly agree

Whales should have the legal right to bodily liberty (the right not to be imprisoned)
Strongly disagree  Disagree  Neither agree nor disagree  Agree  Strongly agree

Great apes should have some legal rights like the right to be free from harm
Strongly disagree  Disagree  Neither agree nor disagree  Agree  Strongly agree

Some elephants are more deserving of the right to bodily integrity than some fish
Strongly disagree  Disagree  Neither agree nor disagree  Agree  Strongly agree

Dolphins should not be kept as pets because they are autonomous beings
Strongly disagree  Disagree  Neither agree nor disagree  Agree  Strongly agree

Biomedical testing on mice should be permitted
Strongly disagree  Disagree  Neither agree nor disagree  Agree  Strongly agree

Biomedical testing on great apes should be allowed
Strongly disagree  Disagree  Neither agree nor disagree  Agree  Strongly agree

I would represent a chimpanzee’s right to bodily liberty (the right not to be imprisoned) if it were threatened
Strongly disagree  Disagree  Neither agree nor disagree  Agree  Strongly agree

Whales should have the legal right to a life free from harm
Strongly disagree  Disagree  Neither agree nor disagree  Agree  Strongly agree

Elephants should have the legal right to be free from imprisonment
Strongly disagree  Disagree  Neither agree nor disagree  Agree  Strongly agree

End of Block: Experimental Group C

Start of Block: Demographic Information

What is your age?

In what year of law school are you?
 o  1L
 o  2L
 o  3L

Where do you attend law school?
 ▼ DROP DOWN MENU

Which term do you use to describe your gender identity?
 o  Man
 o  Woman
 o  Non-binary
 o  Other
 o  Prefer not to answer

Are you color blind?
 o  Yes
 o  No

What is your favorite color of the options below?
 o  Red
 o  Orange
 o  Yellow
 o  Green
 o  Blue
 o  Purple
 o  Pink

Do you intend to pursue a career related to environmental law? If no, please list the area of law you intend to pursue.
 o  Yes
 o  No
 o  Unsure
APPENDIX B

Email to law schools regarding survey distribution
Subject: brief request re undergrad thesis project
Dear [insert title/name],
My name is Elza Bouhassira and I am senior in the Sustainable Development program at Columbia University. I am conducting a survey as an essential part of my undergraduate senior thesis. Law students are the target population of my survey and due to your role as [insert title] I am reaching out to see if you would be willing to send the brief description of my survey below and a link to my survey to your current student body.
Description: This study aims to understand law students’ opinions concerning the idea that nature should have certain legal rights. The survey takes about 4-7 minutes and consists of 3 parts. At the end of the survey, participants will have the option to enter their e-mails to be put into a raffle for one of five $20 Amazon Gift Cards.
[INSERT LINK TO SURVEY]
Thank you so much for your time!
Best wishes,
Elza Bouhassira
BA Candidate in Sustainable Development
Columbia University

APPENDIX C

Email to raffle winners
Subject: thesis survey raffle winner!
Dear [insert name],
Last month you took a few minutes to fill out my senior thesis survey on law students’ opinions on creating legal rights for nature. I’m writing to let you know that you won one of the five $20 Amazon gift cards that were being raffled off for participation in the survey!
Attached is a Print at Home gift card from Amazon worth $20.
Thank you so much for participating in my research!

Best wishes,
Elza Bouhassira
BA Candidate in Sustainable Development
Columbia University

APPENDIX D

Survey pilot testing questions
[Insert Interview Number]
Say to participant: I am testing the survey that will be used in my senior thesis. I am giving you minimal instructions right now because the participants who take the survey will only have what is written. If any questions do not concern you because they ask about law school, feel free to make up a response or skip the question.
Post-Survey Completion Questions:
• Did you feel at any point that the survey was too long?
• Were any questions confusing?
• Did you look at the photographs?
• Were there enough instructions?
• Were the instructions clear?
• Are there any other concerns you have or anything you feel could be improved in the survey design?
[List how long the survey took the participant to complete]

APPENDIX E

Expert interview script
For the eight interviews, I based the conversation around the scripts below, using one if the interviewee was a lawyer and the other if they were a photographer. I started the interview by explaining my project and answering any questions the interviewee might have about what my research looked at. Depending on the prior knowledge of the interviewee, I skipped one of the questions. For example, when interviewing Steven Wise, I did not ask about his familiarity with the NhRP approach. Given that the interviews were semi-structured, the script was a jumping-off point and sometimes the conversation wandered.

Interview script:
Started each interview by explaining my project: My senior thesis looks at whether photographs can evoke more positive responses toward the idea that nature and nonhuman animals should have certain legal rights like the right to exist or the right to be free from imprisonment. Three approaches are being studied overall, but two of the three are the main focus. The first is called Earth jurisprudence and aims to shift the foundations of modern governance systems to be more in line with a world view that sees humans as one part of a greater Earth system. The second is an approach used by the nonprofit the Nonhuman Rights Project which aims to get recognition of some nonhuman animals’ rights to freedom from imprisonment and the freedom to not be experimented on by obtaining a writ
of habeas corpus that recognizes a nonhuman animal was wrongfully detained by its owners. The third is the argument that animals most similar to humans are most deserving of rights. That’s a pretty brief summary, is there anything I should elaborate on or that you have questions about before I ask you some questions I have prepared?

For lawyers:

• Were you previously familiar with either of these strains of thought? Have you done any work related to either?
• In your opinion, is giving nature and nonhuman animals some legal rights an effective way to address environmental issues? If not, what legal avenue do you think is more effective?
• Can you see Earth jurisprudence ideas impacting the work you do on a day-to-day basis?
• The third argument that is being looked at is that nonhuman animals most similar to humans deserve some legal rights more than nonhuman animals which are dissimilar. Do you believe the similarity approach is a necessary first step toward a greater shift in the system, if one were to occur?
• Do you think it is possible that the idea to give nature and nonhuman animals some legal rights could become mainstream in legal scholarship?
• Are there any statutes or court cases you think are particularly relevant? Is there anyone else you think I could talk to or whose research I could look into?

For the photographer:

• Do you think negative or positive images have a stronger impact on people’s behavior? Why?
• Have you ever had a specific image or series that really got people’s attention? What about those visuals do you think was particularly striking?
• In your experience how do people react when they see a disturbing image of an animal being mistreated?
• My survey hopes to see if photographs can evoke more positive responses to granting nature and nonhuman animals some legal rights, do you know of any similar work or anyone else whose work I could look into?

Bibliography


Rickard, Laura N., Jonathon P. Schuldt, Gina M. Eosco, Clifford W. Scherdt, and Ricardo A. Dazia-no. “The Proof is in the Picture: The Influence...


EXPLORING THE RISE AND FALL OF BITCOIN PRICES: TO WHAT EXTENT WAS OPTIMISTIC HERDING BEHAVIOR IN BITCOIN MARKETS DISRUPTED BY THE INTRODUCTION OF BITCOIN FUTURES CONTRACTS?

MANAS RAJASAGI, PENN STATE UNIVERSITY CLASS OF 2019

ABSTRACT

Bitcoin is a novel technology and global economies are in the early stages of incorporating cryptocurrencies into their existing economic systems. Given the limited understanding of Bitcoin from a policy and economic perspective, this paper examines the extent to which futures trading accentuated Bitcoin’s price collapse. This analysis is done through the lens of behavioral economics, specifically regarding herding, or the propensity of investors to disregard their private information and emulate the decisions of other investors.

A key finding from this analysis reveals that herding occurred following the introduction of Bitcoin futures contracts when the market was experiencing extreme negative daily returns. This paper offers economic context for policy makers seeking to better understand the effects of derivative markets on Bitcoin’s prices by way of the behavioral characteristics of its participants.

ACKNOWLEDGMENTS

I would like to thank Dr. Karl Schurter for his time and willingness to share his expertise to guide me through my research and analysis. I would also like to thank Dr. James Tybout for his continuous feedback on my work. I extend my deepest gratitude to the larger Penn State Economics faculty for their commitment to excellent teaching, and for the conversations over the years that have continuously inspired me to explore economics and my interests beyond it.

I am incredibly grateful to my parents and friends for their unconditional support and seemingly endless words of encouragement throughout my undergraduate career. I would also like to thank Maham Quraishi for helping me see the bigger picture, and pushing me to grow through challenging times.

At the time of this writing, there were 17,583,587 Bitcoins in circulation, with a spot price of $3871.18 USD, leading to Bitcoin’s market capitalization of $68,069,278,684 USD. Bitcoin’s rapid gain in market valuation is an affirmation that it has captured the minds of futurists, entrepreneurs, scientists, revolutionaries, and ordinary citizens alike. Cryptocurrencies seek to disrupt, or end, the underlying systems governing financial services, market transaction protocols, and monetary policy tools. Bitcoin is the most popular instance of such a technology.

The successful integration of cryptocurrency technologies into global economic systems requires a responsible adaptation of their use, guided by appropriate and balanced regulation. Although Bitcoin’s function is to evade regulation and intervention by bodies such as the SEC, legislative policy makers, Federal Reserve, etc., it is naïve to assume that governments and traditional financial institutions will standby without interference. The echoes of the 2008 Financial Crisis serve as a reminder to regulators that no innovation in financial products or services that attracts significant public investment and euphoria should go unmonitored and unchecked. Given Bitcoin’s ability to grow its market capitalization so quickly, regulators will likely continue to play a pivotal role in Bitcoin’s journey, and regulation will inevitably affect Bitcoin’s price and derivatives offerings over time. However, Bitcoin regulation is complex; the technology is not yet fully understood, and the ramifications of its mass adaptation are not clearly defined. For example, some consider Bitcoin a commodity, while others view it as a currency, and some are of the opinion
that it is a new asset class in and of itself. If cryptocurrencies are to replace traditional currencies, then it is critical to understand the intricacies of its markets and the behavioral characteristics of its stakeholders (e.g. investors, users, optimists, and pessimists). Herein lies the intent of this paper, which is to further our understanding of the behavioral tendencies of Bitcoin's stakeholders in response to regulations, specifically with respect to the potential introduction of derivatives. This paper accomplishes this by using an empirical methodology for measuring herd behavior in markets from Christie and Huang (1995), and interpreting the model results through the lens of established theoretical and empirical literature from behavioral economics.

This paper is motivated by an economic letter composed by Hale et al. (2018) of the Federal Reserve Bank of San Francisco (FRBSF). Hale et al. (2018) propose a correlation between the introduction of futures contracts by the Chicago Mercantile Exchange and the steep fall of Bitcoin's prices. Hale et al. (2018) refer to the trend in Figure 1, where it is clear that Bitcoin was experiencing a rapid rise in prices between 2017 and 2018, but prices began a downward spiral following the introduction of futures contracts on December 17, 2018 on the Chicago Mercantile Exchange. They plot the performance of the S&P index to highlight the abnormality of Bitcoin's price movements in comparison to a traditional market indicator. To the authors of Hale and al. (2018), Bitcoin's sudden decline is not a surprising result, since previous literature has revealed the effects that short selling instruments, in this case futures, could have on an asset's underlying price.

Although Hale et al. (2018) liken Bitcoin’s trajectory to the introduction of credit default swaps to the housing market and its subsequent crash, they do not explore Bitcoin’s price decline from the perspective of herding. This paper aims to fill that void and add to their work on the possible association between Bitcoin’s decline and the introduction of futures trading. The results of this analysis could prove useful for future regulations on derivatives offerings for Bitcoin.

LITERATURE REVIEW
RELEVANT THEORY FROM BEHAVIORAL ECONOMICS & FINANCE
A foundational understanding of relevant principles from behavioral economics and finance is critical to parsing the implications of the experimental results presented later in this paper. Considering Bitcoin’s major price correction, it is attractive to classify its rapid growth and decline as a bubble; however, Bitcoin’s story is more complicated. Brunnermeier (2001) identifies the occurrence of a bubble as when an asset’s price is inflated above its fundamental value, and its price is determined endogenously (Brunnermeier, 2001). As Hale et al. (2018) concede, it is difficult to determine Bitcoin’s fundamental value. However, Bitcoin’s fundamental value could be broken down into its transactional value and its speculative value. Transactional demand is driven by the actual use of Bitcoin to purchase goods and services, while speculative demand is driven by investors trading Bitcoin in order to sell their currency for profit in future time.
periods (Hale et al., 2018). Bitcoin’s novel technology inhibits the currency’s ability to quickly adapt for use in mainstream services and goods transactions. Thus, it is far easier for speculative demand to remain active as investors may continue predicting future increases in value regardless of Bitcoin’s utility to an ordinary consumer. Bitcoin was a proving ground for distributed and decentralized currency; however, its future transactional value depends on the number of users on its network willing and able to pay the transaction fees necessary for Bitcoin miners to maintain the network (Huberman, Leshno, Moallemi, 2017). Given these points, the possibility that Bitcoin’s price will once again break its peak price cannot be fully discredited.

Nevertheless, Bitcoin’s steep price growth and decline is a reflection of investors’ beliefs and market decisions, which could be described as either rational or irrational. Rational investors, according to Barbaris and Thaler (2002), must satisfy two checks: they must correctly update their beliefs as new information is discovered, and their choices must be consistent with Savage’s notion of Subjective Expected Utility (SEU); meaning they must choose between options based on their expected payoffs and the probability of them occurring. The authors position rational behavioral finance as an alternate explanation to the irrational behavior seen in financial markets, which traditional theory cannot fully account for (Barbaris and Thaler, 2002). Traditional financial theory points to the efficient markets hypothesis (EMH), which states that asset prices immediately reflect all available and new information (Malkiel, 2003). Under this traditional view, investors are all rational and equally aware of an asset’s true value.

While the EMH completely discounts the possibility of bubble formation, Abreu and Brunnermeier (2003) present a model in which bubbles can persist even with rational investors. Setting the context for their model, Abreu and Brunnermeier (2003) acknowledge that there are certainly a limited number of agents subject to irrationality in markets; however, the authors propose a model focused on rational arbitrageurs (i.e. participants who aim to profit from market inefficiencies). They argue that rational arbitrageurs aim to maximize their returns by riding the bubble and form a diverse set of expectations on optimal timing for exiting the market. The resulting variance in exit strategies and market timing accelerates bubble growth, as no single arbitrageur wants to exit the market too early and risk sub-optimal returns. A relevant outcome from the model is that large scale synchronized action taken by the arbitrageurs is necessary to burst a bubble; however, they lack the incentive to make such a coordination effort if prices are continuously rising. With no incentives for joint action, except for possible synchronization events like economic shocks (e.g. commodity shortages, natural disasters, political shifts), the bubble will persist despite arbitrageurs’ knowledge that the asset is over-valued.

![Figure 2: Formation of Bubbles with Rational Arbitrageurs](image)

The model in Figure 2 is Abreu and Brunnermeier (2003)’s illustration of bubble persistence starting at a time t₀ and ending at t, where some exogenous event pops the bubble. The time t₀+η indicates the point when all traders have knowledge of the bubble’s existence. The authors model investors’ optimism by setting the growth rate of price, g, in pₜ=e^gt, to be greater than the risk-free interest rate, r (Abreu and Brunnermeier, 2003). This model is very relevant to Bitcoin’s own price history, since optimistic sentiment in the value of new technology could not be checked by pessimistic pressure until the introduction of regulated Bitcoin futures. The introduction of Bitcoin futures could be a potential instance of a synchronizing exogenous event that leads arbitrageurs to coordinate their actions in a sell-off.

Barberis and Thaler (2002) set forth three limits to arbitrage, which can supply some additional context to the Abreu and Brunnermeier (2003) model. The first limit is the fundamental risk of a decrease in asset value, which arbitrageurs typically balance by shorting a substitute asset. The second limit is noise trader risk, where an asset price decrease can be further pressured down by unsophisticated investors, causing arbitrageurs to liquidate their positions. The third limit is implementation costs, which include transaction
costs on trades, short sale limitations, and information acquisition costs to identify mispricing (Barberis and Thaler, 2002). In the context of Bitcoin, a trader may balance fundamental risk by diversifying across multiple cryptocurrencies. This may be challenging in practice for unsophisticated investors seeking portfolio diversification, as differences amongst cryptocurrencies can be difficult to parse without technical knowledge. A large number of unsophisticated and irrational investors would lead to higher noise trading, which in turn results in high volatility and unpredictability (De Long, Shleifer, Summers, & Waldmann, 1990). De Long, Shleifer, Summers, & Waldmann (1990) further explain that the unpredictability of irrational investors creates additional risk for arbitrageurs since the initial mispricing of the asset may become increasingly severe over time.

A central idea to this analysis is the formation of herds, and the impacts they have on asset prices. In an early work, Banerjee (1992) describes herds as everyday occurrences where we decide to follow the lead of other peoples’ decisions. Banerjee’s model classifies agents who pay attention to the decisions of others as rational, since these agents are attempting to incorporate information known to others but unknown to themselves. He discovers that agents inadvertently form herds in an effort to incorporate others’ decisions, all while ignoring their own private information. As a result, any given herding agent’s information is less valuable to other agents because it does not reflect their own private information (Banerjee, 2002).

FUTURES & FINANCIAL INNOVATION IN MARKETS

The key event of interest to this study is the introduction of futures trading for Bitcoin. Relevant literature on the impacts of futures trading on the price movements of their underlying assets in other sectors can offer insights into Bitcoin’s price behavior. Futures contracts are financial derivatives that allow a buyer and seller to engage in the exchange of the underlying asset at a pre-determined time, price, quantity, and quality (Chen, 2019). The contracts themselves are standardized according to a set of rules determined by the issuing authority and are legally binding. On the first day of Bitcoin futures trading at the CME, more than $100 million in notional value was generated through the trading of more than 1000 contracts across all available maturities (McCourt, 2017). Figure 3 below shows the significant increase in the magnitude of futures trading in the 312 trading days leading up to March 15, 2019, relative to when they were first introduced in mid-December, 2017. The data show notional values reaching above $400 million on three occasions during this period, with the majority of other peaks averaging around $200 million. Figure 3 also shows the notional value of open interest during this period, where open interest is defined as the total number of contracts in existence that have not yet been liquidated (Grant, 2019). The key takeaway from the data is that the Bitcoin futures market is highly active, and the increase in the aggregate notional value of contracts shows that investors are becoming more comfortable with engaging in futures transactions. For more information on CME Bitcoin futures contract specifications and standards, see Appendix C.

Investors mainly purchase futures with the intention to decrease risk by hedging their positions on the underlying asset, or to profit from speculating on the asset’s future price. It is important to note that speculators do not need to hold any position on the underlying asset itself. As Hale et al. (2018) explain, speculators with pessimistic views on Bitcoin can purchase futures contracts listing Bitcoin’s price lower than the current spot price. Upon expiration, contracts are financially settled through cash (please see Appendix C for more information on CME settlement procedures). Prior to the introduction of futures contracts on the CME, there was no such regulated and mainstream mechanism to express pessimistic sentiment on Bitcoin via market transactions. Herein lies their significance to Bitcoin.

Hale et al. (2018) also reference findings from Fostel and Geanakoplos (2012), who showed how financial innovation in the form of Credit Default
Swaps (CDS) allowed pessimistic investors to bet against the underlying mortgage-backed securities (MBS) that grew to such immense popularity in the run-up to the 2008 housing bubble. The authors’ data show that tranching and seemingly endless securitization of MBS led to a sustained increase in their prices, while the introduction and popularization of CDS destroyed the market. They further explain that if the introduction of CDS had been timed earlier in the period of rising securitization, then the asset bubble would not have grown as large as it did, and the subsequent crash would not have been as severe (Fostel and Geanakoplos, 2012).

The discussion in Fostel and Geanakoplos (2012) has interesting parallels to Bitcoin’s own demise; however, there are several differences to note. For one, there was no regulated securitization of Bitcoin in the period leading up to its price increases, rather there was mainly optimistic speculative demand from people willing to purchase it through Bitcoin’s own network infrastructure. Additionally, the size of the Bitcoin market was nowhere near as large as the housing market when the housing bubble burst in 2008. Furthermore, due to Bitcoin’s unregulated nature, it attracted minimal institutional investment prior to the introduction of futures contracts, although this may have changed ex-post. As the Chicago Mercantile Exchange noted, “Commodity Trading Advisors (CTAs), ETF providers, hedge funds, sell-side firms, proprietary trading firms and retail traders, among others all voiced strong interest in a bitcoin futures contract” (McCourt, 2017). As both the theoretical literature discussed previously and the forthcoming empirical literature will show, institutional investors are observed to behave differently than individual investors in markets. This point is important to retain for the discussion of this paper’s findings later on.

RELEVANT EMPIRICAL LITERATURE
Numerous empirical analyses were conducted on various sections of global financial markets to detect and measure herding. The empirical methodology used in this paper follows that proposed by Christie and Huang (1995), and is discussed in detail in the research methodology section. In brief summary, Christie and Huang (1995) propose using the cross-sectional standard deviation (CSSD) to determine the propensity of individual equity returns to stray from market returns during periods of large price changes. Herding implies minimal propensity to stray, while large CSSD values indicate no herd consistent behavior. Christie and Huang (1995) define the 1% and 5% tails of the returns distribution as extreme swings in portfolio returns.

Interestingly, Christie and Huang (1995) found that there was little evidence of herding during periods of market stress as indicated by relatively positive and high values of CSSD. Rather, their results were more in line with the predictions of rational asset pricing models, which suggest an increase in CSSD between a security’s returns and average portfolio returns during periods of market stress (Christie and Huang, 1995). Additionally, the authors found lower CSSD values when returns were in the extreme 5% tails of market returns than in the extreme 1% tails. This indicates that rational asset pricing theory is more pronounced at the extreme 1% tails than at the 5% tails of market returns. Christie and Huang (1995) also extend their returns analysis from daily data to monthly data so as to not neglect herd formation or dissolution over longer time intervals. They find that CSSD values are higher for monthly price data than they are for daily data, and attribute this result to the increased time available for investors to make decisions, resulting in a higher likelihood of deviation of equity returns from the portfolio average (i.e. more dispersion and less herding). For the authors, this observation conveys that herding is more of a short-lived phenomenon.

Hwang and Salmon (2004), contrary to Christie and Huang (1995), assert that confining herding to periods of market stress assumes that herds cannot form during seemingly normal market scenarios. They cite a case of NASDAQ, which represents a portion of the stock market, experiencing large price movements, while the overall stock market did not reflect any notable changes. Hwang and Salmon (2004) argue that definitions of market stress are also subjective, which could result in varied results across different implementations of Christie and Huang’s (1995) method. Additionally, they believe it is important to distill “spurious herding,” which results from collective actions taken by investors based on changes in fundamentals, from the herding resulting from investors suppressing private information (Bikchandani and Sharma, 2001). To address this concern, Hwang and Salmon (2004) condition their model on changes in an asset’s fundamentals, allowing them to distill spurious herding from overall herding observed in returns data. Additionally, Hwang and Salmon (2004) warn that the possible correlation between CSSD and time series
volatility makes it difficult to distinguish herd effects from volatility effects (Hwang and Salmon, 2004). Applying their method to the US and South Korean stock markets, they found evidence of herding during both rising and falling markets. Notably, they assert that market crises actually serve as correctional events that drive market prices towards their efficient levels (Hwang and Salmon, 2004). Similarly, the authors attribute the tendency of herds to weaken before a market crisis manifests and pursue a “flight to fundamentals” to Christie and Huang’s (1995) inability to detect herds during periods of peak market stress, i.e. the 1% tails of Christie and Huang’s (1995) returns distribution. Essentially, Hwang and Salmon (2004) find that herding decreases before a market downturn (for more information on Hwang and Salmon’s (2004) methodology see Appendix A).

Chang, Cheng, and Khorana (1999) expand Christie and Huang’s methodology by three measures: they use the Cross Sectional Absolute Deviation (CSAD) with a non-linear returns parameter, differentiate herding between developed and developing markets, and analyze changes in herding following the liberalization of Asian markets. Including a non-linear returns parameter can reveal additional interaction between the CSAD and market returns otherwise overlooked in the linear case, i.e. herding may be increasing at a decreasing rate or vice versa. Finally, the authors construct their analyses for both a bull market and bear market case in order to detect any asymmetry in herding between those two scenarios. Using the CSAD, they found no herd consistent behavior in developed financial markets, namely the U.S., Hong Kong, and Japan, which is consistent with Christie and Huang’s (1995) results. However, significant non-linear herding was detected in developing markets, specifically South Korea and Taiwan, which the authors attribute to investors basing their decisions on macroeconomic factors in lieu of complete firm specific information in these countries. The latter point is significant; nascent markets lacking complete and equally distributed information seem to cultivate herd behavior. Notably, CSAD values increased at a higher rate in an improving market than they did during declining markets, which indicates that herding decreased at a higher rate in up-markets than during down markets (see Appendix A for more information on Chang, Cheng, and Khorana’s (1999) methodology).

Caparrelli, D’Arcangelis, and Cassuto (2004) apply the methods from Christie and Huang (1995), Hwang and Salmon (2004), and Chang, Cheng, and Khorana (1999) to the Italian stock market. The authors studied price movements of 68 large-cap stocks and 83 small-cap stocks over a 13-year period. Using Christie and Huang’s (1995) CSSD method, they found no evidence of herding during extreme market conditions (1%, 5%, 17% upper and lower tails of the returns distribution), which is consistent with Christie and Huang’s (1995) observations. Next, the authors reformulated dispersion using the CSAD during both the bull-market and bear-market case, as prescribed by Chang, Cheng, and Khorana (1999). In this case, they found positive and significant coefficients for the linear parameter for both the bull-market and bear-market cases, where the linear bull-market coefficients were larger than their bear market counterparts. This indicates higher dispersion during rising markets than during falling markets. However, if the non-linear parameter is negative and significant, then we can say that herding is present, since the CSAD is increasing at a decreasing rate as market returns become more extreme. The authors do find negative and significant coefficients for the non-linear parameter, which indicates herding in large-cap and small-cap stocks, as well as the global portfolio consisting of both large-cap and small-cap stocks. Finally, using Hwang and Salmon’s (2004) approach, they found that herding is subject to significant variation over time, and the extent of herding is typically larger for small-cap stocks than for large-cap stocks. Herding tends to form more gradually in the small-cap subsample, while it develops more rapidly in the large-cap sample. They note the increase in herding in the Italian stock market beginning in December of 1998 of their sample, which has some consistency with Hwang and Salmon’s (2004) finding that herding typically peaks at the end of bull markets and decreases before an imminent crash.

Tan et al. (2008) follow Chang, Cheng, and Khorana (1999), using CSAD to detect herd behavior in the Chinese stock market. As the authors explain, the Chinese Stock market is divided into two classes: A shares, which allow for individual investors to purchase stocks, and B shares, which can only be purchased by institutional investors. This is a valuable opportunity for research, since individual investors are generally less knowledgeable and more capital restricted than institutional investors. The authors exploit the difference in share class to distill differences in herd formation between the two groups, which can help outline key differences in their respective behav-
ioral tendencies. Their key findings include that CSAD values tend to be larger for extended time intervals, such as monthly data, and are lower for weekly and daily data. Note that this is consistent with Chang and Huang (1995), where CSSD values were lower for daily returns data than for monthly returns data. Furthermore, CSAD values are notably higher for class B shares, which are purchased by institutional investors, than for class A shares, which are purchased by individual investors. The authors note that this could be attributed to the informational advantage that institutional investors have over individual investors, which allows them to stay somewhat ahead of the curve and reallocate their assets. Meanwhile, the individual investors may be far more sensitive to signals from others’ trading behavior, and respond based on their observations of such information.

BITCOIN AND BLOCKCHAIN TECHNOLOGY
A brief note on how Bitcoin and Blockchain technology function is important in clarifying the important distinctions between Bitcoin, fiat currencies, and equities. As outlined in the white paper by Nakamoto (2008), Bitcoin is a self-regulating, peer-to-peer transaction system that eliminates the need for financial intermediaries and counter-parties, shifting complete control and anonymity to its users. The Bitcoin network is comprised of a globally distributed set of nodes, also known as miners, which expend time and energy to crack cryptographic puzzles, called proof-of-work, to mine more Bitcoin. Nodes are incentivized to mine Bitcoin by being awarded small amounts of the cryptocurrency when they solve cryptographic puzzles, as well as through transaction fees. The puzzles must be solved in order for a transaction to be processed and added to the transaction chain. The network is virtually secure from attack since copies of the transaction ledger are distributed and stored remotely on all nodes in the system, and the majority nodes have to agree on one version of the transaction chain. The likelihood that an attacker could race an honest network node and modify the Blockchain decreases exponentially as the length of the Blockchain increases (Nakamoto, 2008). In effect, as transactions are processed and added to the existing chain, the network actually becomes more secure. Bitcoin achieves decentralization, privacy, and security using Blockchain technology. A model of Bitcoin’s Blockchain is shown in Figure 4. Figure 4 illustrates how transactions are sequentially processed and linked together in a chain by nodes on the network. Each user has a unique private key, a randomly generated number, and a corresponding public key.

![Figure 4: Bitcoin Blockchain Model](https://via.placeholder.com/150)

A user’s public key is essential to validate that they indeed initiated a transaction or are the recipients of a transaction. Suppose a user wishes to purchase an item and sends the seller some amount of Bitcoin. The sender uses his private key to sign off and encrypt the transaction, which is broadcasted to the network. The only way to decrypt the sender’s transaction is to validate that the public key corresponds to its private key, which is the work of the network nodes (Berentsen and Schar, 2018).

It is clear that Bitcoin offers many favorable characteristics as a means of exchange; however, there is no shortage of skepticism about its adaptation. As Huberman et al. (2017) show, without sufficient congestion, i.e. transaction volume on the Bitcoin network, nodes’ expected return on processing transactions decreases, resulting in their departure from the network. As the number of active nodes decreases, the network security decreases, and users will be less willing to transact, which threatens to further decrease transaction volume. The latter exemplifies a technical challenge for maintaining Bitcoin’s long-term success. Numerous public figures in government and financial services have denounced Bitcoin, and expressed a lack of confidence in its ability to effectively address the requirements of scalable currency. Notably, the Reserve Bank of India, India’s central bank, has banned its use, citing that there are various risks associated in dealing with virtual currencies (Reserve Bank of India, 2018).
Undoubtedly, government regulation and opinions expressed by officials and leaders in industry will impact Bitcoin’s price. The essential point is that this market is not nearly as ordered in terms of clarity of information and future regulation as are other markets, like conventional fiat currencies. To this point, there is still much debate on the classification of Bitcoin into an existing asset class, like currencies or commodities. An investigation of Bitcoin’s public ledger by Baur et al. (2018) shows that a third of Bitcoin is held by investors who never use it to transact for goods and services, while a much smaller segment of users actually uses it as they would use a traditional currency. The authors have several classifications for how they define an investor versus a normal user (see Appendix C).

RESEARCH METHODOLOGY
CHRISTIE AND HUANG’S APPROACH TO DETECTING HERDING

This paper follows Christie and Huang (1995)'s method of using the Cross-Sectional Standard Deviation (CSSD), also referred to as dispersion, as a proxy measure to detect changes in investor herding during periods of market stress, i.e. time instances with abnormally high or low price movements. Following literature from behavioral economics and social psychology, Christie and Huang (1995) define herding as an investor’s tendency to suppress his/her private beliefs and make decisions based on observations of others’ actions. In theory, as Christie and Huang (1995) postulate, the result of such behavior should manifest itself in the form of a market consensus, where the deviation between individual returns and overall average market returns should be relatively small. Determining the CSSD for each time interval (e.g. daily, weekly, monthly) in the time series is a simple and intuitive method for quantifying how an individual asset’s returns move relative to the average returns of the portfolio.

Christie and Huang (1995) believe that herds are most likely to form during periods of abnormal market conditions, which for the purpose of their analysis were defined as extreme price movements of the equities included in their portfolio. The fundamental idea is that during periods of market stress, investors enter a panic mode and adopt a belief that the market must have information they themselves do not, leading them to condition their decisions on observations of other investors. Likewise, when the market is undergoing abnormal price movements, Christie and Huang (1995) expect the herd mentality to proliferate, resulting in lower dispersions. The authors concede that their hypothesis is contrary to the predictions of rational asset pricing theory. Under rational asset pricing models, extreme market events result in increased dispersion between individual equities and the market/portfolio’s returns since each security has a unique sensitivity to market events (Christie and Huang, 1995). Christie and Huang (1995) define dispersion using the following formula, which again, is simply the cross-sectional standard deviation of individual equity returns relative to the portfolio at time t.

\[
S_t = \sqrt{\frac{\sum_{i=1}^{n}(r_{it} - \bar{r}_t)^2}{n-1}},
\]

, where \((i)\) indexes an equity in the portfolio, and \((t)\) indexes time.

To detect herd formation during periods of market stress, Christie and Huang (1995) regress dispersion on extreme market returns. The effects of extreme market returns on dispersion are isolated through dummy variables for abnormally high or low returns, which are classified as either the top or bottom 1% or 5% of the returns distribution. The following describes their model.

\[
S_t = \alpha + \beta_1 D_{1t} + \beta_2 D_{2t} + \epsilon_t
\]

\(\alpha\) Represents the average dispersion in portfolio returns between the two extreme tails
\(D_{1t} = 1\) If portfolio returns on day \(t\) are in the lower tail of the returns distribution
\(D_{1t} = 0\) Otherwise
\(D_{2t} = 1\) If portfolio returns on day \(t\) are in the upper tail of the returns distribution
\(D_{2t} = 0\) Otherwise

Negative values for \(\beta_1\) and \(\beta_2\) imply \(S_t\) decreases when returns are extreme, which is indicative of herding since dispersion between individual equity returns and market returns is decreasing. Rational asset pricing would expect \(\beta_1\) and \(\beta_2\) to be positive, since it predicts that dispersion should actually increase during periods of market stress.
Visually, as shown in Figure 5, Christie and Huang (1995) are interested in capturing the effects of returns in the extreme tails (1% and 5% criterion) of the returns distribution on dispersion.

In context of Christie and Huang’s (1995) hypothesis, the regression in equation (1) should produce negative coefficients for both dummy variables. This would indicate a decrease in dispersion, i.e. an increase in herding, when price returns are abnormally low or high. Positive coefficients would reflect the opposite case, which supports the rational asset pricing theory that dispersion increases, i.e. herding decreases, with abnormal price movements. Lastly, it could also be the case that herding manifests itself asymmetrically, such as when one coefficient is negative and another is positive. This could happen when herding increases during extreme negative price movements (i.e. $\beta_1$ is negative), while decreasing during extreme positive price movements (i.e. $\beta_2$ is positive), or vice versa. Christie and Huang (1995) further investigate the asymmetrical case by comparing their estimates for dispersion with those predicted by a rational asset pricing model. This step is beyond the scope of this analysis; however, it is interesting to note that Christie and Huang (1995) found convincing evidence in favor of rational asset pricing as the magnitude of market downturns increases. In any case, the magnitude and directionality of the herding coefficients are the most important output from this analysis, and reveal key insights about investor behavior during market stress.

ADAPTATION OF CHRISTIE AND HUANG’S METHOD

Although for the purposes of their analysis Christie and Huang (1995) use price data on conventional equities, their methodology can offer valuable tools to identify herd behavior in the cryptocurrency market. Limitations caused by this adaptation are discussed in the next section.

Similar to Christie and Huang (1995), this paper is interested in isolating the effects of extreme price movements on herding in the Bitcoin market. In addition, this paper also seeks to isolate the effects of herding before and after the introduction of Bitcoin futures on December 17, 2017, which coincides with Bitcoin’s peak price. Isolating the effects of extreme market returns before and after futures will reveal differences, if any, in investor herding during Bitcoin’s bullish phase and its subsequent bearish crash phase. Again, this is made possible by the subsequent decline of Bitcoin’s value following the introduction of futures contracts.

These regressions are conducted individually on three portfolios: a pure cryptocurrency portfolio, a pure fiat currency portfolio, and a mixed portfolio consisting of both cryptocurrencies and fiat currencies. The purpose of constructing multiple portfolios is to identify any possible links between investor behavior in fiat currency markets and cryptocurrency markets, and to make an effort to quantify any relationship between the two. Additionally, applying this methodology to a conventional and well understood asset like fiat currencies offers some level of standardization, since cryptocurrencies are relatively nascent. Likewise, since the same regression is run on all three portfolios, the only changes in model coefficients should be driven by the nature of the respective market being investigated.

Additionally, Christie and Huang (1995) find significant differences and valuable insights from running their regressions on both daily and monthly price data on the same portfolio, since herd behavior is shown to change as the period of observation changes. Since cryptocurrencies are relatively young, monthly price data is limited, resulting in numerous regression issues and limiting test power. An additional issue encountered was multi-collinearity resulting from limited observations of the dummy variables following the introduction of futures. To solve for this and still achieve variation in time intervals, daily and weekly price data are used instead. Multi-collinearity was addressed by redefining the upper 1% tail following the introduction of futures using the price data specific to that period.

The regression of dispersion on market returns during periods of stress is identical to that outlined previously and in Christie and Huang (1995). However, the regression used to test for changes in herding before and after the introduction of futures is different and involves two additional dummy interaction
variables that allow for time-splitting. This is noted in equation (2) on the next page.

\[
S_t = \alpha + \beta_1 D_{t1} + \beta_2 D_{t2} + \beta_3 (D_{t1} \times \text{Time}) + \beta_4 (D_{t2} \times \text{Time}) + \epsilon_t,
\]

where:

- \( S_t \) is the average dispersion portfolio in returns between the two extreme tails.
- \( D_{t1} = 1 \) if portfolio returns on day \( t \) are in the lower tail of the returns distribution.
- \( D_{t2} = 0 \) Otherwise.
- \( D_{t2} = 1 \) if portfolio returns on day \( t \) are in the upper tail of the returns distribution.
- \( D_{t2} = 0 \) Otherwise.
- \( \text{Time} = 1 \) if the time, \( t \), \( t \geq \text{December 17, 2017} \).
- \( \text{Time} = 0 \) Otherwise.

This regression is only applied to the daily and weekly data for the pure cryptocurrency portfolio and the mixed cryptocurrency-fiat portfolio, since the futures contracts are specific to Bitcoin.

**ASSEMBLY OF DATASET**

As outlined in the regression methodology, the primary goal of this analysis is to use the cross-sectional standard deviation as a proxy measure to detect herding. Given the nature of this regression, it is important to maximize the number of observations so as to allow for a sufficient number of instances where the market is experiencing extreme positive and negative swings in asset returns. In addition to sufficient time series observations, it is also valuable to apply the regression to a standard portfolio of fiat currencies. Lastly, a mixed portfolio consisting of both cryptocurrencies and fiat currencies can help to control for any interactions between the two markets, despite their fundamental drivers being very different.

The pure cryptocurrency portfolio consists of daily price data for 19 cryptocurrencies chosen based on their market capitalizations. Price data were collected from coinmarketcap.com, which also provided information on market capitalizations. The cryptocurrencies are not weighted by their share of overall portfolio market capitalization. The cryptocurrencies included in the portfolio are Bitcoin (BTC), Ripple (XRP), Ethereum (ETH), Bitcoin Cash (BCH), Stellar (XLM), EOS (EOS), Litecoin (LTC), Tether (USDT), Cardano (ADA), Monero (XMR), Dash (DASH), NXT (NXT), Doge Coin (DOGE), Tron (TRON), IOTA (IOTA), NEM (NEM), NEO (NEO), Ethereum Classic (ETC), and ZCash (ZEC). It is assumed that a high market capitalization indicates investor and public confidence in the cryptocurrency, and their willingness to use it either for transactional or speculative purposes. Bitcoin is the oldest of these currencies, and so it has the most available price data. In order to justify the cross-sectional standard deviation, only price data following April 1, 2015 were used in the regressions. Around this time, most of the cryptocurrencies in the portfolio had been introduced, and thus can be used to calculate more convincing values of the CSSD.

The pure fiat currency portfolio is mainly constructed of globally dominant currencies, denominated in U.S. dollars, from high GDP countries; however, some low profile currencies were also included as an attempt to balance the portfolio and better reflect global price movements. Data were collected from the Federal Reserve Bank of St. Louis FRED exchange rate data. The currencies in the portfolio are the U.S. Dollar (USD), Japanese Yen (YED), British Pound (GBP), Swiss Franc (CHF), Canadian Dollar (CAD), New Zealand Dollar (NZD), Chinese Yuan (RMB), Indian Rupee (INR), Brazilian Real (BRL), and Thai Bhat (THB). Although data were available since as far back as the year 2000 between all of the currencies in the fiat portfolio, the regression was also restricted to time periods after April 1, 2015. This constraint maintains consistency with the time period used for the pure cryptocurrency portfolio.

The mixed fiat-cryptocurrency portfolio is a merged set of all the currencies included in both the pure cryptocurrency portfolio and the pure fiat currency portfolio. The purpose of creating a merged portfolio is to allow for regression outputs to show any effects that the fiat currency markets could have had on cryptocurrency markets. For example, since many cryptocurrencies are denominated in dollars and other major global currencies, a significant change in the price of a US dollar could affect the relative price of Bitcoin. Again, the regression was restricted to time periods after April 1, 2015.

All raw price data across the three portfolios were collected as daily prices. Before applying the regression model, weekly price data were generated using STATA by finding prices at the beginning and end of 7 day periods. Weekly data were then stored separately for analysis. For both daily and weekly price data, returns were calculated using the following formula.
LIMITATIONS OF AVAILABLE DATA AND REGRESSIONS

There are some key limitations of both the data used in this analysis and the regression models applied to understand it. The principal distinction between this analysis and that of Christie and Huang (1995) is that the latter focused specifically on equities and this analysis focuses on cryptocurrencies. Equities, which represent ownership of small portions of companies, bear significant differences from currencies. One difference is that investors purchase equities for dividends and voting rights, while a currency actually allows for transactions between holders for goods and services. Moreover, this analysis is interested in cryptocurrencies, which themselves have several fundamental differences from conventional currencies. As mentioned earlier, Baur et al. (2018) find that a significant portion of Bitcoin is held by investors, not those using it for transactions, which differentiates it from a classical currency.

Furthermore, compared to Christie and Huang (1995), who use 26 years of equity price data for their daily NYSE and Amex data and 63 years of monthly data for NYSE firms, this analysis is restricted to a little more than 3 years of data. The consequence of a more limited time series within the context of Christie and Huang’s (1995) methodology is fewer observations of periods of market stress (i.e. extreme high and low returns). Another result of the limited data was multiple collinearity issues between the explanatory variables, which were periods of market stress, on CSSD. This was solved for by restricting the data in equation (2) to the 5% and 10% extremes in order to collect enough observations of extreme price movements. As mentioned by Hwang and Salmon (2004), the definitions of market stress are very subjective and can change over time as a market develops. This is important to keep in mind while comparing the results of this analysis to those discussed in the empirical literature review which also apply Christie and Huang’s method. Another important point that Hwang and Salmon (2004) consider is incorporating the effects of changes in fundamentals on herding. They incorporate the dividend price ratio, relative treasury bill rate, term spread, and the default spread to account for macro-economic impacts. This is difficult to emulate for Bitcoin since its fundamentals are difficult to estimate, and also because attempts to link its performance to macro-economic factors would be shallow due to the limited understanding we have to date on this relationship.

Some final effects on the results not captured could be new regulations, speculation in news media on Bitcoin, hacks of cryptocurrency exchanges, etc. These variables are not controlled for, and so can have possible explanatory impact on returns behavior throughout the time series.

HERDING ANALYSIS RESULTS

DISPERSION AND MARKET STRESS MODEL RESULTS

Table 1: Dispersion Regressed on Market Stress (Daily Price Data)

<table>
<thead>
<tr>
<th>Panel A: Pure Cryptocurrency Portfolio</th>
<th>1% Criterion</th>
<th>5% Criterion</th>
<th>10% Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model Variable</td>
<td>a</td>
<td>b</td>
<td>c</td>
</tr>
<tr>
<td>Coefficient</td>
<td>.523257</td>
<td>.027272</td>
<td>.531702</td>
</tr>
<tr>
<td>t-statistic</td>
<td>42.93</td>
<td>1.77</td>
<td>19.99</td>
</tr>
<tr>
<td>P &gt;</td>
<td>t</td>
<td>0.000</td>
<td>0.077</td>
</tr>
</tbody>
</table>

Panel B: Pure Fiat Currency Portfolio | 1% Criterion | 5% Criterion | 10% Criterion |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Model Variable</td>
<td>a</td>
<td>b</td>
<td>c</td>
</tr>
<tr>
<td>Coefficient</td>
<td>.006479</td>
<td>.017764</td>
<td>.013591</td>
</tr>
<tr>
<td>t-statistic</td>
<td>17.80</td>
<td>11.12</td>
<td>16.20</td>
</tr>
<tr>
<td>P &gt;</td>
<td>t</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Panel C: Mixed Cryptocurrency and Fiat Currency | 1% Criterion | 5% Criterion | 10% Criterion |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Model Variable</td>
<td>a</td>
<td>b</td>
<td>c</td>
</tr>
<tr>
<td>Coefficient</td>
<td>.045735</td>
<td>.0459546</td>
<td>.2854408</td>
</tr>
<tr>
<td>t-statistic</td>
<td>44.93</td>
<td>4.55</td>
<td>29.54</td>
</tr>
<tr>
<td>P &gt;</td>
<td>t</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Notes: Regression coefficients are specific to equation (1) $\alpha + \beta_1 D_E + \beta_2 D_T + \epsilon$, see regression methodology section for additional information on model parameters.
upward or downward movements in market returns, while the 5% and 10% are relatively less extreme.

Beginning with Panel A in Table 1, which consists solely of cryptocurrencies, results show that all of the regression coefficients are positive across the 1%, 5%, and 10% criteria. Aside from the beta coefficients for the lower tail of the 1% and 5% criteria, all other coefficients are significant at the 95% confidence level. Notably, the coefficients for the upper tails of the distribution are all larger than those representing the lower tails. In all cases, the positive coefficients show that dispersions are increasing during periods of market stress; however, dispersion tends to increase much more significantly when the returns are positive and extreme, i.e. the upper tails. For example, the ratio of the increase in dispersion during extreme positive returns to extreme negative returns in nearly 13:1 for the 1% criterion, and falls to around 12:1 and 7:1 for the 5% and 10% criteria respectively. Lastly, a comparison of the coefficients between the three criteria show that dispersions increase as the market returns become more extreme.

Panel B studies the impacts on dispersion for the portfolio of fiat currencies. The results are similar to those form Panel A, and all of the coefficients are significant at the 95% confidence level. In comparison to their corresponding values in Panel A, the coefficients in panel B are much smaller, indicating that the magnitude of increases in dispersion for fiat currencies are smaller than for cryptocurrencies. An additional comparison of the constants in the regressions between Panel A and B show that, on average, dispersion is much lower in the fiat currency portfolio than the cryptocurrency portfolio.

Panel C addresses the interaction between cryptocurrencies and fiat currencies through a joint portfolio of returns data. All coefficients are positive and significant at the 95% confidence level across the three market criteria. The magnitude of coefficients is far more comparable to that seen in the cryptocurrency portfolio than in the fiat currency portfolio. As also observed in Panel A and Panel B coefficients, dispersions in Panel C are the highest for extreme returns (i.e. at the 1% criterion) and decrease sequentially for the 5% and 10% criteria. Observations from the weekly data, as shown in Table 2 below, reveal important differences that arise from changing the observation interval. Referring first to Panel A, dispersion coefficients across the lower tails of the 1%, 5%, and 10% returns distributions are negative, while the coefficients on the upper tail are positive. Negative dispersion coefficients indicate the presence of herd behavior; however, the coefficients for the lower tails are not statistically significant at the 95% level, which decreases confidence in indicating herd behavior during extreme negative returns for weekly data. Additionally, statistically significant dispersion coefficients are larger than they were for daily data in Panel A from Table 1, indicating that dispersion is larger in week-to-week observations than day-to-day time intervals.

Panel B from Table 2 is very consistent with Panel B from Table 1. All dispersion coefficients are positive and significant at the 95% confidence interval across 1%, 5%, and 10% criteria. Additionally, as observed in Panel B from Table 1, dispersions tend to increase as the market undergoes more extreme price movements. Coefficients are again larger for the weekly returns data than they were for daily data in Panel B from Table 1.

Interestingly, Panel C from Table 2 consists of all positive coefficients on dispersion as well. Since Panel C data is a combination of both cryptocurrencies and fiat currencies, the effect of conjoining the datasets removed the herding observed in the lower tails of the cryptocurrency portfolio presented in Panel A. Additionally, dispersions are higher, as observed earlier, during periods of extreme positive returns in prices as opposed to when they are extreme negative returns. All coefficients, excluding the ones corresponding to the lower tail of the returns distribution are statistically significant at the 95% confidence interval.
DISPERSION AND MARKET STRESS BEFORE AND AFTER INTRODUCING FUTURES MODEL RESULTS

The next segment of results is derived from equation (2), which distinguishes herding during periods of market stress before and after the introduction of Bitcoin futures contracts. Table 3 below displays regression coefficients from daily returns data, while Table 4 shows the corresponding coefficients for weekly data. In comparison to Table 1 and Table 2, there are only two panels in Table 3 and Table 4 since the futures contracts were specific to Bitcoin and not the larger currency market. Please refer to the methodology section for additional details.

Table 3: Dispersion Regressed on Market Stress Before and After Futures Introduction (Daily Price Data)

<table>
<thead>
<tr>
<th>Panel A: Pure Cryptocurrency Portfolio</th>
<th>1% Criterion</th>
<th>5% Criterion</th>
<th>10% Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model Variable</td>
<td>α</td>
<td>β</td>
<td>γ</td>
</tr>
<tr>
<td>Coefficient</td>
<td>.09317</td>
<td>.0100301</td>
<td>.029229</td>
</tr>
<tr>
<td>t-statistic</td>
<td>13.02</td>
<td>3.33</td>
<td>10.79</td>
</tr>
<tr>
<td>P &gt; [t]</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Mixed Cryptocurrency and Fiat Currency</th>
<th>1% Criterion</th>
<th>5% Criterion</th>
<th>10% Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model Variable</td>
<td>α</td>
<td>β</td>
<td>γ</td>
</tr>
<tr>
<td>Coefficient</td>
<td>.1349299</td>
<td>.0493981</td>
<td>1.252943</td>
</tr>
<tr>
<td>t-statistic</td>
<td>16.95</td>
<td>0.70</td>
<td>11.33</td>
</tr>
<tr>
<td>P &gt; [t]</td>
<td>0.000</td>
<td>0.005</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Notes: Regression coefficients are specific to equation (2). \( S_t = \alpha + \beta_1 D_t^1 + \beta_2 D_t^2 + \beta_3 (D_t^1 \times T_i) + \beta_4 (D_t^2 \times T_i) + \epsilon_t \), see regression methodology section for additional information on model parameters.

In reference to Panel A from Table 3, the dispersion coefficients before the introduction of futures are positive for both the lower and upper tails of the 5% and 10% criteria, while dispersion coefficients after the introduction of futures are negative. While all coefficients are significant at the 95% confidence interval, negative coefficients for the weekly data under normal market stress were not statistically significant. Negative dispersion coefficients indicate the presence of herd consistent behavior following the introduction of futures contracts trading, specifically when the market is experiencing extreme negative returns. As was observed in Table 1 and Table 2, the magnitude of dispersion is higher in Panel A of Table 3 for the upper tails. Furthermore, Panel A of Table 3 affirms again that dispersions tend to be higher at the more extreme 5% criteria than at the 10% criteria.

Shifting to Panel B, the negative dispersion coefficients marking the period after the introduction of futures contracts still persist into the mixed portfolio of cryptocurrencies and fiat currencies, in contrast to how the negative dispersion coefficients become obscured in the previous mixed portfolio data in Panel C, Table 2. Except for the dispersion coefficients for the lower tail of the distribution after futures trading began, all other coefficients are statistically significant. The pattern of dispersion being greater for the upper tails of market returns than the lower tails, as well as for the 5% criteria as opposed to the 10% criteria, still persist in Panel B.

Table 4 on the following page is the final re-
gression in this analysis, and displays the results from using weekly data as opposed to daily time periods. An immediate distinction between Table 4, Panel A results and Table 3, Panel A results is that the dispersion is negative for the lower tail of returns distributions before the introduction of futures. This is in addition to the negative post-futures coefficients. This indicates that herding was present in the lower tail prior to futures, and increased in intensity following the introduction of futures. There is an issue of statistical significance in Panel A, where only the constant and the dispersion coefficient corresponding to the upper tail of the returns distribution before futures are significant. This is potentially a reflection of lowered test power resulting from limited dummy variable observations for the lower tail of the distribution.

<table>
<thead>
<tr>
<th>Panel A: Pure Cryptocurrency Portfolio</th>
<th>5% Criterion</th>
<th>10% Criterion</th>
<th>Panel B: Mixed Cryptocurrency and Fiat Portfolio</th>
<th>5% Criterion</th>
<th>10% Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model Variable</td>
<td>Coefficient</td>
<td>t-statistic</td>
<td>P&gt;</td>
<td>t&lt;</td>
<td>Model Variable</td>
</tr>
<tr>
<td>a</td>
<td>.151249</td>
<td>15.69</td>
<td>0.000</td>
<td>a</td>
<td>.1437705</td>
</tr>
<tr>
<td>b</td>
<td>-0.0355</td>
<td>0.72</td>
<td>0.472</td>
<td>b</td>
<td>-0.01891</td>
</tr>
<tr>
<td>c</td>
<td>.018537</td>
<td>13.43</td>
<td>0.000</td>
<td>c</td>
<td>.0184754</td>
</tr>
<tr>
<td>d</td>
<td>-0.027391</td>
<td>-0.34</td>
<td>0.734</td>
<td>d</td>
<td>-0.026441</td>
</tr>
<tr>
<td>e</td>
<td>-0.220564</td>
<td>-1.88</td>
<td>0.064</td>
<td>e</td>
<td>-0.1935682</td>
</tr>
</tbody>
</table>

Notes: Regression coefficients are specific to equation (2): $S_t = a + \beta_1 D_t^5 + \beta_2 D_t^{10} + \beta_3 (D_t^5 \times Time) + \beta_4 (D_t^{10} \times Time) + \epsilon_t$, see regression methodology section for additional information on model parameters.

Table 4: Dispersion Regressed on Market Stress Before and After Futures Introduction (Weekly Price Data)

Looking at Panel B in Table 4, the results are more consistent with those observed in the daily data since the only negative dispersion coefficients are those corresponding to the period after futures were introduced. Again, the negative coefficients indicate that herding increased in the period after futures. Furthermore, the dispersions in the upper tail are larger than those in the lower tail across both the 5% and 10% criteria. Additionally, the dispersion coefficients are also larger for the 5% criteria than they are for the 10% criteria, which reaffirms that dispersions are increasing as the intensity of market stress increases. With the exception of the dispersion coefficients corresponding to the lower tail before futures, and the coefficient for the lower tail after futures, all other coefficients are significant.

INTERPRETATION OF FINDINGS

This study was separated into two empirical analyses; the first analysis measured the effects of extreme returns (i.e. market stress) on herding, and the second analysis measured changes in market stress before and after the introduction of Bitcoin futures contracts. While the first analysis provides important context for herding in the Bitcoin market during market stress, the second analysis narrows the scope to specifically determine the impact of futures on investor herding.

The results of the first analysis have many similarities to results from Christie and Huang (1995). With respect to the first analysis, herding was not detected during periods of market stress for the cryptocurrency, fiat currency, and mixed cryptocurrency-fiat currency portfolios. Furthermore, individual cryptocurrency returns actually deviated from their market portfolio to a much greater extent during periods of extreme positive market movements, consistent with the rational Efficient Market Hypothesis. The primary conclusion from these results is that herding is not a concern during extreme price movements for Bitcoin. The failure to detect herd behavior at extreme conditions may be an implication of Huang and Salmon’s (2004) criticism, where the authors detect herding during periods not characterized by market stress. A comparison between the daily and weekly data reveals that dispersion increases as the time interval increases. As Christie and Huang (1995) postulate, this could be a result of investors incorporating more information into their decisions, the effects of which materialize over longer periods. This result is consistent with results observed in Caparelli et al. (2004) and Tan et al. (2008), who also detect higher dispersions as the time interval increases. Essentially, investors are less likely to herd as the time period increases. As noted earlier, there is an indication in weekly returns data that investors are herding during extreme negative price movements. This coefficient is not statistically significant, possibly because there are not enough instances of extreme negative returns during this period to justify this relationship. However, it is noteworthy.
The results from the second analysis are more nuanced and important to answering the question about whether herding was the cause of Bitcoin’s steep decline. Across both time daily and weekly data, the introduction of futures decreased dispersion for both the lower and upper tails of the returns distribution. Notably, herding was detected following the introduction of futures contracts for the pure cryptocurrency portfolio. This applied to both the 5% and 10% criteria. It is important to note that herding was only detected in the lower tail and not the upper tail of the returns distribution following futures trading. This suggests investors have a higher propensity to herd, following the introduction of futures, when the market is experiencing large negative price movements. Herding was no longer detected for the mixed portfolio case. With respect to weekly returns data, the model indicates that investors exhibited herd behavior when market returns were negative before the introduction of futures, and herding further increased after they were introduced. The issue is that these coefficients do not satisfy the 95% confidence level, however, the repeated trend towards herding during extreme negative price movements is interesting to note. Again, herding was not observed in the mixed portfolio, although it is clear that dispersions decreased following futures trading.

The daily data from Table 3 provide the most significant results to this analysis, both in terms of statistical significance, as well as in terms of offering a strong explanation of investor behavior in response to futures. Table 3 results showed that herding was present in daily returns data when returns were extremely negative. With respect to theory, initial price devaluations driven by futures trading from pessimists could have been exacerbated by investors’ tendency to herd when market returns are extremely negative. In relation to Abreu and Brunnermeier’s (2003) model about bubble formation, the evidence from this study suggests that futures could have been the exogenous event at time that began Bitcoin’s price decline. Taking from Banerjee’s (2002) claim that herd behavior is rational, and setting that in the context of Abreu and Brunnermeier (2003), it is reasonable to conclude that rational herding could have accelerated Bitcoin’s bear market following the introduction of futures contracts. As Banerjee (2002) also notes, as investors continuously incorporate others’ decisions into their own decisions, the inherent value of this information decreases over time, which is threatening to overall market health.

The challenge is that there should have been some level of herding observed during the period leading to its peak value, however the data do not indicate herding before futures. In effect, there could have been non-herd related market characteristics of investor decision making that led to Bitcoin’s steep price increase. What this analysis can conclude is that herding was present when daily returns were extremely negative following the introduction of futures contracts. It is also significant that dispersions consistently increased when market returns were positive and extreme, which is in line with the rational asset pricing hypothesis. This is not necessarily a result of asymmetric distribution of information, since we should expect a poor information landscape to be defined by more herding, as Chang et al. (1999) observe in developing Asian markets.

Given the results of this study, it is not conclusive that herding was the fundamental driver of Bitcoin’s steep price increases, although herding was observed for negative daily returns after the introduction of futures. This conclusion is made in the context of the limitations of Christie and Huang’s (1995) methodology, as discussed in the limitations section. It is possible that herding could be observed to a greater extent when controlling for Bitcoin’s fundamentals, such as changes to its technology or macroeconomic events, and during periods not characterized by market stress. It is also possible that herding could be more apparent when Bitcoin’s time series is divided into a bull market phase and a bear market phase, as Caparrelli et al. (2004) do for the Italian stock market.

Further research should incorporate the effects of transaction volume on Bitcoin’s network, as well as price volatility to study how these varied before and after futures trading began. Additionally, possible implications from news events and Google search-term popularity can serve as proxy measures of Bitcoin’s popularity and propensity to attract investment. Lastly, it would be valuable to see how the mix of institutional and individual investors changed before and after the introduction of futures. While futures contracts allow greater participation from institutional investors, it would be interesting to model how institutional investment in actual Bitcoin, as opposed to derivatives, would affect the market. This investment is currently limited due to Bitcoin’s short history, limited market maturity, and developing regulatory landscape, among other factors.
The authors developed a new measure of herding using cross-sectional dispersions that incorporate factor sensitivities of the equities being studied. Their approach is also able to discern herding caused by changes to the equities fundamentals from herding driven by the suppression of investors’ own beliefs. They specifically study the US and South Korean stock markets. The herding parameter used by Hwang and Salmon (2004) is summarized and simplified in Caparelli et al. (2004), and is shown below.

\[
H(m, t) = \text{var} \left( \frac{\beta_{it} - 1}{\sqrt{s^2_{it}}} \right)
\]

Where:

- \(H(m, t)\) represents the herding parameter
- \(\beta_{it}\) represents the stock \(i\)'s beta at time \(t\), where \(\beta_{it}\) is also classified as a parameter given by a Capital Asset Pricing Model (CAPM) in equilibrium. The CAPM is shown below:

\[
e_{it} = \beta_{it}e_{m} + \epsilon_{it}
\]

\(s^2_{i}\) represents variance in the beta of the stock

\(s^2_{m}\) represents variance in market

Low values of \(H(m, t)\) indicate the presence of herding, while higher values are far less convincing about the presence of herd behavior. The information above is summarized from both Hwang and Salmon (2004) and Caparelli et al. (2004), please refer to their papers for more in-depth discussion of methodology.

**APPENDIX B**

**DISTRIBUTION OF RETURNS**
APPENDIX C
FUTURES CONTRACTS AND ADDITIONAL TERMS SPECIFICATIONS

CME BITCOIN FUTURES CONTRACTS SPECIFICATIONS (AS OF 3/31/2019)

<table>
<thead>
<tr>
<th>Contract Unit</th>
<th>5 bitcoin, as defined by the CME CF Bitcoin Reference Rate (BBR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price Quotation</td>
<td>U.S. dollars and cents per bitcoin</td>
</tr>
<tr>
<td>Trading Hours</td>
<td>Sunday - Friday 6:00 p.m. - 5:00 p.m. (5:00 p.m. - 4:00 p.m./ CT) with a 60-minute break each day beginning at 5:00 p.m. (4:00 p.m./CT)</td>
</tr>
<tr>
<td>Minimum Price Fluctuation</td>
<td>Outright: 5.00 per bitcoin ~ $25.00 Calendar Spread: 1.00 per bitcoin = $5.00</td>
</tr>
<tr>
<td>Listed Contracts</td>
<td>Quarterly contracts (Mar, Jun, Sep, Dec) listed for 2 consecutive quarters and the nearest 2 serial months.</td>
</tr>
<tr>
<td>Settlement Method</td>
<td>Financially Settled</td>
</tr>
<tr>
<td>Termination Of Trading</td>
<td>Trading terminates at 4:00 p.m. London time on the last Friday of the contract month. If that day is not a business day in both the U.K. and the U.S., trading terminates on the preceding day that is a business day for both the U.K. and the U.S.</td>
</tr>
</tbody>
</table>

This information is copied directly from the CME website, please refer to their paper for more clarification on definitions and additional information on their methodology.

Classification of Inventories by Baur, Hong, and Lee

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active investor</td>
<td>More than two transactions and only sends Bitcoin in greater than USD$2000 transactions.</td>
</tr>
<tr>
<td>Passive investor</td>
<td>More than two transactions and only receives Bitcoin in greater than USD$100 transaction with no sending of Bitcoin, or has made only one receiving Bitcoin transaction of greater than USD$100.</td>
</tr>
<tr>
<td>Currency user</td>
<td>Makes more than one transaction, has made both sending and receiving transactions and sending transaction sizes are less than USD$2000.</td>
</tr>
<tr>
<td>Tester</td>
<td>Makes only one transaction of less than USD$100.</td>
</tr>
<tr>
<td>Miner</td>
<td>A user that has ever mined for Bitcoin as identified by a user receiving newly generated Bitcoin.</td>
</tr>
<tr>
<td>Hybrid user</td>
<td>All other users not categorized.</td>
</tr>
</tbody>
</table>

For more information on settlement procedures, position limits, exchange rules, etc., please refer to the CME Group’s website, www.cmegroup.com.

CME Bitcoin Futures Daily Settlement Procedure (As of 3/31/2019)

CME Bitcoin Futures Daily Settlement Procedure (As of 3/31/2019)

Normal Daily Settlement Procedure

Tier 1: Trades on CME Globex

All contract months settle to the volume-weighted average price (VWAP) of outright trades between 14:59:00 and 15:00:00 Central Time, the settlement period.

Tier 2: CME Globex Market Data

In the absence of trades during the settlement period, the contract month settles to the midpoint of the Bid/Ask between 14:59:00 and 15:00:00 Central Time, the settlement period.

Tier 3: Absence of Two-Sided Markets

If there are no two-sided markets available during the settlement period in the Lead* month, then the settlement price will be the last trade price (or prior settle in the absence of a last trade price) adjusted to the Bid/Ask if one side is present. Deferred months will settle to the net change of the previous month, adjusted to the Bid/Ask if one side is present.

*The Lead month is the contract expected to be the most active.

Final Settlement Procedure

Delivery is by cash settlement by reference to the Final Settlement Price, equal to the CME CF Bitcoin Reference Rate (BBR) on the Last Day of Trading.

Bibliography


THE HELVIDIUS GROUP
COLUMBIA UNIVERSITY