ACKNOWLEDGMENTS

The Helvidius Group thanks the following institutions and individuals for their guidance and financial support:

INSTITUTIONAL SPONSORS
Activities Board at Columbia
Weatherhead East Asian Institute, Columbia University
Institute of Latin American Studies, Columbia University

ACADEMIC AND ADMINISTRATIVE ADVISORS
Professor Van C. Tran, Faculty Advisor
Veronica Baran, Administrative Advisor

BENEFACTORS AND DONORS
Peter Tomassi

EDITORS EMERITI
CONTENTS

INTRODUCTION
1 Editor's Note: The View from South Lawn
   Peter Rutkowski

GUEST ESSAY
3 A Modern Tragedy? COVID-19 and U.S.–China Relations
   Professor Thomas J. Christensen

ESSAYS
9 Gender, Party, and the Press: Media Coverage of the Second “Year of the Woman”
   Julia Press

36 An Ethical Evaluation of the Use of Predictive Risk Models in Health and Human Services: A Case Study on the Allegheny Family Screening Tool
   Samuel Oh

56 Grappling With the Datafied Self: College Students and Wearable Fitness Trackers
   Meghan McMahon

89 The Buck Passer's Dilemma: Evidence from Japan
   Matthew Salah

109 Tribal Affiliation or Class-Based Differentiation: Class, Tribal Identity, and Interpretations of State-Led Tourism Development in Petra, Jordan
   Nicolas Reeves

130 Death From Above: Evaluating the Impact of Drone Strikes on AQAP Activity
   Chris Conrad

157 Beyond Mano Dura: An Integrated Security Strategy for the Northern Triangle
   Ryan Witmer Keating
There’s no doubt that the fall of 2020 was one of the most highly anticipated periods in recent American history. An increasingly divided United States, in many ways still reeling from the outcome of the 2016 Presidential Election, and still in the depths of the COVID-19 crisis, approached Tuesday, November 3rd, 2020 full of anticipation and anxiety, and with, arguably, a sense of resolution and purpose not experienced for decades.

This sense of purpose stemmed in no small part from a new generation of voters, which had been either too young or insufficiently mobilized to cast a ballot in 2016. Many of this new generation were inspired to make their voices heard, supporting equal rights for minority communities, fighting for justice, and challenging long-standing structures and institutions.

This spirit drove much of the scholarship in our Fall 2020 edition. While American college students were largely exiled from their campuses, the Journal of Politics & Society received not only a larger pool of submissions than during any previous semester (260+), but submissions of a quality easily up to par with semesters past. With so many outstanding submissions, the Executive Board made the decision to increase the number of publications to eight articles, an exciting development that speaks to the vibrancy of undergraduate research in the social sciences.

First we present an article by Julia Press, which focuses on the 2018 election cycle, labeled the “Year of the Woman.” Press analyzes more than 700 media sources to find that while coverage of female candidates increased over the 2018 cycle, unique challenges continue to plague female candidates of both parties. While on the presidential debate stage Donald Trump and Joe Biden wrangled over the growth of technology’s power in the United States, Meghan McMahon used her Colgate University campus as a testing ground, investigating the impact of fitness trackers on student health behaviors, emotions, and agency. Her article not only spoke to our editorial board, but is a first for the Journal as it researches at the intersection of technology and socialization.

Stepping away from the national spotlight, we turn to author Samuel Oh’s case study on the ethics of predictive risk assessment tools in health and human services. Oh analyzes child protective services tools—specifically the Allegheny Family Screening Tool used in Allegheny County, Pennsylvania—and disputes claims that the tool is founded on a basis for wrongful discrimination.

One of this edition’s most unique articles was also local in focus, but this time international. Nicholas Reeves contributed an ethnographic study of Petra, Jordan, drawing-out the important implications of state-sponsored tourism on the region’s tribal communities.

Any edition would not be complete without discussion of important topics in international relations and foreign policy. In “The Buck-Passer’s dilemma,” Matthew Salah uses evidence from modern Japan to argue that many US allies remain militarily weak for institutional reasons beyond their international security relationships. Also in this section, Chris Conrad’s investigation of the impact of the United States’ Obama-era drone campaign in Yemen on the activities of Al-Qaeda in the Arabian Peninsula (AQAP) offers important perspective as the United States continues its extensive use of drones in the region. Finally, the Fall 2020 edition concludes with a focus on Latin America, with Ryan Keating’s deep analysis of security strategy in the Northern Triangle (of Guatemala, Honduras, and El Salvador).

As made apparent from the discussion above, the scope and quality of our published articles remains expansive and first-rate, despite the challenges to research, writing, editing, and publishing posed by
the pandemic both here in New York and across the United States.

Most importantly, I want to recognize the entire Fall 2020 editorial board and the wonderful leadership I had supporting me throughout the entire selection and editing process; these individuals made the transition to a virtual semester as painless as it could possibly be, and speak true to the Journal’s mission of promoting excellence in undergraduate research, no matter how challenging times may be.

Peter Rutkowski

*Editor in Chief*
The term “tragedy” is not generally used in contemporary discourse the way that it was originally intended in literature and theater. Nowadays, anything awful is labeled a “tragedy”: from the criminal—such as a mass shooting—to the accidental—such as a train wreck or bus accident. In ancient Greek theater, the outcomes of tragedies were usually very bad as well, but it was not just the negative outcome that differentiated tragedies from comedies. A tragedy requires that the characteristics of the key players—often termed “tragic flaws”—and their interactions in a particular context, lead them unwittingly, but not simply accidentally, to a crescendo of pain and suffering. In some instinctive sense, the tragic outcome seems avoidable. But puzzlingly, it also seems inevitable once one drills into the personalities and flaws of the key players and the context within which they interact.

It is in this sense of the word that I see COVID-19 as a potential tragedy in U.S.–China relations and a potential tragedy for the world. The nature of the two countries’ political realities in 2020 have led to mismanagement of the crisis on both sides of the Pacific. The two sides have so far squandered historic opportunities for cooperation against a common enemy, including in their interactions with other actors, such as the World Health Organization (WHO). And other than the metaphorical invasion of earth by space aliens, we are unlikely to meet a more common enemy in our lifetimes than this coronavirus. If great powers, including strategic competitors like the United States and China, cannot cooperate on countering this threat to the entire globe, then how can we expect to cooperate on other issues?

So why has COVID-19 so far proven to be a source of greater friction, rather than greater cooperation, between the People’s Republic of China (PRC) and the United States of America? Here is where the tragic flaws of the two political systems collide. In China, where the epidemic began, it seems that the government handled the original outbreak of the virus very poorly. The local governments in the city of Wuhan and the surrounding Hubei province apparently suppressed the bad news that a virus was spreading in the city, silencing through coercion the voices of doctors who were blowing whistles and pointing to the dangers of an epidemic. Until January 20th of this year, the Chinese government did not even recognize publicly that the disease was clearly being passed between humans. But the disease has proven itself so contagious in multiple countries, it seems impossible to believe that health care workers in Wuhan were not among the early patients, which would be a very clear sign of human-to-human transmission. After all, with full knowledge of the contagion, and very careful practices in place today to fight the spread in hospitals in the United States, many health care workers and hospital staff have still contracted the disease while treating COVID-19 patients.

At a minimum, then, it seems that there was a local cover-up of the dangers of the disease. Chi-
nese nationals I spoke with in the United States were fully aware and not surprised that concerned doctors in Wuhan were stifled by strict regulations against “spreading rumors” or “revealing secrets” without prior permission from higher authorities. The lack of a free press in China also hampered the prompt dissemination of knowledge about the disease to the general public in Wuhan and beyond. Local officials’ reluctance to draw attention to problems is predictable in a system that blames and often punishes such officials for bad outcomes, even if forces that were generally outside of their control were the cause. Added to the mix is the massive anti-corruption drive launched by current President Xi Jinping. Most officials in a broadly corrupt political system fear being selected for a “disciplinary investigation” that has only one predictable end. If there is no independent justice system and most officials have at least some black marks in their dossiers, then local officials try to avoid missteps like taking maverick, public positions on negative occurrences such as disease spread. For these reasons, local officials sweeping bad news under the carpet and the early and quite consequential paralysis in responding to something like COVID-19 should not come as a surprise to scholars of contemporary Chinese politics.

Most likely there were also further cover-ups at higher levels in the Chinese Communist Party (CCP). Central government elites do not want to see the PRC’s reputation tarnished on the international stage, and more importantly, want to insure that the CCP’s legitimacy at home is not harmed by coverage of the origins of the pandemic and the weak and even destructive early responses to it. And there is real reason for the CCP leadership to worry. The local problems that I describe above are not really local, but rather the natural outcomes of a single-party authoritarian state without institutions, like a free media and an independent court system, that could protect the individual rights of citizens, including whistleblowers, against state repression. Chinese nationals in the United States with whom I discussed these issues in late January and February did not know what to think as the crisis escalated, but some expressed fears for their families back home based on one shared perception: the CCP could not be trusted to tell the public the truth about what was actually happening.

Once the central government recognized the spread of the contagious disease and locked down Wuhan on January 23rd, the Chinese government appears to have been quite effective at limiting the spread of the disease, rapidly expanding hospital capacity, distributing protective gear to health care workers, expanding testing protocols, and isolating, often forcibly, those with proven disease and even those suspected of having been exposed to the disease. Chinese doctors and health care officials almost certainly have learned valuable lessons to share with the outside world, including the United States. This is true even if it is the case that the same system in which they work caused tremendous damage early on by allowing a large, international city like Wuhan to become a giant incubator for a highly contagious and dangerous disease that would spread through the country and around the world. The reluctance of the World Health Organization to label COVID-19 a global health emergency until the end of January, a full week after the lockdown of a large, international Chinese city, may also have caused significant damage. While later investigations will likely reveal more fully why this delay occurred, it does seem probable that what was at work was some combination of Chinese political pressure on WHO member states or WHO’s top leadership to preserve the PRC’s reputation on the international stage and the WHO’s overreliance on official reports from member states like China.

While the WHO’s slowness in coming to that conclusion may have delayed reactions to the coming catastrophe in various parts of the world in consequential ways, oddly the one place that this does not seem to have been the case is the United States. Andironically, it has been Washington that has become the loudest critic of the organization. Here is where the American tragic flaws come into play. During the Trump administration, the U.S. government has downgraded the importance of science and expertise in its decision-making processes and, under the banner “America First,” has generally avoided using multilateral organizations and agreements to protect and assert U.S. interests. Under President Trump, fewer government health experts were on the ground in the U.S. mission in China than in past administrations. Deep expertise combined with long government experience is associated in top administration political circles with the so-called deep state that Trump has accused of trying to undermine his presidency. The president himself clearly prefers making decisions based on
his gut instincts and on his hopes rather than on the results of careful research. So he claimed early on that the disease posed limited risk to the American society or economy; that it would disappear soon “like a miracle,” perhaps when there was warm weather; and, more recently, that injecting disinfectants could be explored as a potential cure for the disease.

Even after the disease appeared threatening to the world, including the United States, the Trump administration largely dismissed the danger as overblown or, worse, inflated intentionally by the administration’s political opponents. It is very difficult to see how an earlier recognition by the WHO that the novel coronavirus constituted a global health emergency would have changed that flawed American reality.

Critically important weeks were lost in implementing serious policies to combat the disease. And the much touted ban on travel from China at the end of January, however sensible, apparently did little to stem the tide of the disease in the United States, since it had already arrived earlier and had begun spreading. In the case of New York, the disease apparently arrived indirectly from China via Europe, before the travel bans on both regions were established. Subsequent repeated claims that tests were universally available and were being provided in sufficient numbers to meet the challenge were patently untrue, and, early on in the crisis, governors were often left to fend for themselves and compete with one another in acquiring protective gear and medical equipment for physicians, in some important cases from China.

When the history of COVID-19 is written, South Korea, New Zealand, and Taiwan will likely be seen as the best examples of free societies that wrestled effectively with the virus in its early phases. Unfortunately, the United States almost certainly will not. For its part, Beijing’s international reputation has apparently taken a big hit not just in the United States but in Europe and the Asia Pacific because of the issues raised above. Despite some impressive reactions in China after the epidemic was publicly recognized and despite efforts to assist other countries with medical equipment and expertise—and thereby boost China’s international image—it appears that COVID-19 will prove much more of a liability than an asset in the PRC’s diplomatic portfolio.

And here is where the systemic insecurities of the Chinese Communist Party and the political and psychological insecurities of the Trump administration seem to be playing off each other in a classically tragic manner. The tragedy is evident in Chinese Foreign Ministry spokesperson Zhao Lijian forwarding conspiracy theories about the U.S. Army planting the disease in Wuhan (he has not been fired and there is no evidence that he was even disciplined for this outrageous accusation). The CCP has also named all new cases of the disease in China as foreign, leading to discrimination against foreigners around the country, especially African migrants in the area in and around Guangzhou. This poor treatment of Africans who have lived in the area for years has done severe harm to China’s reputation in Africa, which had been fostered over many years through economic interaction and infrastructure investment.

The tragedy is also present in President Trump’s race-baiting description of the disease as the “Chinese virus,” once he began taking it seriously, and his apparent celebration on Twitter of a corrected increase in the Chinese official death totals. One can only surmise that he saw the reports of increased Chinese suffering as good because it might somehow make his own government look somewhat more effective in comparison. The president has subsequently stopped using the “Chinese virus” label, but leaked guidance to Republican senators on how to respond to questions about COVID-19 in a presidential election year instructs them to blame China and the WHO for all problems in the United States and to praise Trump for limiting travel from China. Such deflection and scapegoating might impress the president’s political base at home but will almost certainly further harm the United States’ reputation on the international stage and make future cooperation with China more difficult.

Potentially compounding and catalyzing the tragedy, the Biden campaign has decided that attacking Trump for being too soft on China is the best way to leverage the pandemic for political gain in the 2020 elections. This cynical move is akin to Hillary Clinton’s shift to economic nationalism and rejection of the Trans Pacific Partnership in her 2016 campaign. It appears patently disingenuous and destined to derail the U.S. national interest, even if it seems like “good politics” in the toxic domestic political climate in the United States today.

So now we have the makings of a tragedy full of characters with tragic faults: one...
domestically insecure government (the CCP) with a legitimacy crisis that plays out in the form of domestic repression under a single authoritarian leader who can never be questioned or criticized; one insecure administration (that of the U.S.) under a vainglorious president who is running for reelection and demanding that his partisan troops never criticize his response to the COVID-19 crisis, but instead blame China and the WHO for all problems in the country; and one traumatized opposition party (the Democrats), who still can’t believe Trump won in 2016 and who have decided this time around that “when they go low, we should go even lower.” While China was clearly to blame for the crisis in this tragic narrative, so was Trump for being Beijing’s patsy. One criticism that Biden supporters have raised is that early in the crisis Trump permitted the sending of “our [protective] masks” to China. This puts down a marker for the president that any future Sino-American cooperation on the virus in this election year might be called out by the Democrats as somehow traitorous.

I believe that all of the characters in this tragedy would like the virus to go away, but they have all chosen to protect their own reputations by placing blame squarely on others in ways that make much needed international cooperation to combat the virus more difficult. A good dose of self-criticism on all sides will be needed to improve future responses to similar challenges (which will almost certainly arise). More urgently, a good dose of humility and self-reflection might allow for greater international cooperation in this ongoing crisis. Failures of international cooperation will likely cost hundreds of thousands, or more likely millions, of additional lives through disease, hunger, and economic deprivation. Angela Merkel, chancellor of a wealthy and well-equipped Germany, and a responsible and internationally-oriented leader of the first order, recognized recently that we are only in the early stages of this crisis. When one thinks of a much less well-equipped and much poorer sub-Saharan Africa or South Asia, one can only shudder about the number of people who could die later because of fighting now among great power rivals and among the two major political parties in the richest nation on earth.

Whatever mistakes and cover-ups occurred in Wuhan early on, China is now a repository of useful knowledge about the virus and how best to control its spread. It also has a very strong scientific community studying the origins of viruses and their medical treatment, who can cooperate with our own experts both to find a vaccine and to develop effective treatments short of a vaccine. This is true even if it turns out that the virus actually leaked from a scientific facility in Wuhan with insufficient safeguards. There will be time later to assess the early mistakes of China and others in greater detail, but the disease is out there now and we should be tackling it together. And the WHO and other multilateral institutions like the G20 should be bolstered to help address the medical and economic challenges that are likely to spread around the globe, particularly in countries with weak medical infrastructures and poor economies that will almost certainly suffer massive debt defaults. Again this is true, even if international politics and institutional weakness delayed the WHO’s initial response to COVID-19. It simply does not follow any logic (except a tortuous political one) that the proper response to any earlier failures by the WHO should be to cripple the major vehicle of international public health during a global pandemic.

Here are six areas of cooperation that the United States and China can pursue in both bilateral and multilateral settings that would serve their national interests and the interests of the planet, even if they do not necessarily fit the domestic political logics of leaders in Washington and Beijing. The list is suggestive and not intended to be exhaustive and can include cooperation among governments and non-governmental actors.

1. **Share best practices:** The two sides should share and learn best practices for how to slow the spread of the disease, including mistakes to be avoided. While it might be too soon to expect Beijing and Washington to agree to a probe of their early mistakes, it would be very helpful if each side would commit in principle to conduct such a probe after the virus has been brought under control and eliminated. This is unlikely to be our last epidemic. We all need to learn lessons for the long run and it would reduce political tensions between the two nations in the near term to recognize the eventual need for such a probe.

2. **Cooperate on vaccine creation:** The United States and China should work on vaccines together and should pledge to share any breakthroughs with each
other and the rest of the world promptly when they are made. This can be done on a government-to-government basis or in cooperation between universities and companies. One sign of hope on that score is that Chinese and American scientists, including at Columbia University, have managed to perform collaborative research on the disease despite the conflicts between the two governments.

3. Prepare in advance for massive vaccine production and global vaccine distribution: Vaccinating everyone everywhere will be a massive logistical undertaking that will require great forethought before a vaccine is invented. Delays in distribution of even several months could easily cost astounding numbers of lives. If political fighting over who gets vaccines and when were to occur, it would be devastatingly destructive to international cooperation on any matter for years to come.

4. Assist the poorest nations in battling the disease: Cooperate to remediate suffering in the developing world by boosting the medical response capacity in highly vulnerable areas like sub-Saharan Africa, South Asia, etc. In 2014, the United States and China cooperated effectively alongside many other countries to address the Ebola crisis in Africa. And the WHO should be a major actor in this cooperation regardless of any problems related to the organization’s public response in January 2020. And to the degree that the epidemic is accompanied by famines in some places, as seems likely, the United States and China should support the efforts of the World Food Programme to distribute provisions and eliminate distributional bottlenecks slowing the delivery of needed aid.

5. Cooperate to manage debt defaults in the developing world: The possibility of systematic debt defaults in the developing world seems quite real and this could have ripple effects in the entire global financial system. More multilateral cooperation will clearly be needed. The then-brand-new G20 responded rather well to the 2008 financial crisis and should be called upon again to address this global recession. The COVID-19 crisis should also provide an opportunity for global bankers to push China to join international development financing groupings like the Paris Club, which reduce conflicts among lenders when debt crises occur around the globe. Without cooperation on debt restructuring, the international economy could be severely harmed by beggar-thy-neighbor strategies among lending institutions. In this context, the many nontransparent, bilateral infrastructure development loans made by China as part of the Belt and Road Initiative could loom particularly large.

6. Prioritize development of strategic reserves over economic nationalism: Nations are now more acutely aware of their dependence on foreign supplies of needed products in a world of globalization and transnational supply chains. But we should also recognize that global trade has generally been a very positive factor for the world economy and that significant reductions in global trade will likely lead to more, not less, poverty and more, not less, vulnerability to disease and hunger. Two potential solutions to protect global trade would be the diversification of global supply chains so that a single country, like China, is not so essential to supply final manufactured goods. This would mean even more complex economic interactions around the world than we have today, but it would provide a much more efficient solution than each nation trying to produce many products entirely at home to reduce vulnerability. To supplement such a globalist strategy, individual countries should be encouraged to create larger strategic reserves of needed medical and other supplies as an alternative to simply moving all production of such products back to their own countries. Economic nationalism as an alternative to strategic reserves would carry huge opportunity costs for global efficiency and wealth and could also infect international security politics in destabilizing ways. Similar approaches to stockpiling of internationally purchased products for security purposes have long been used effectively in the energy sector.

In order to pursue such a constructive agenda, all countries should call a ceasefire on blaming others over the early outbreak and global spread of the disease. To help facilitate this diplomatic ceasefire, all countries should commit to eventual international investigations into how they responded to the crisis, including mistakes and misdeeds done along the way. The WHO should be involved in such an investigation, and the United States should be actively involved with the WHO to participate and help guide its involvement. For the reasons discussed above, it appears that
neither the PRC nor the United States will likely be pleased to hear the eventual results of such an inquiry. But if they fail to cooperate now and continue to fight, and hundreds of thousands, if not millions, of additional deaths occur as a result, each country will suffer even greater losses to its reputation and diplomatic standing than it would by accepting in advance that it will eventually receive some criticism. The PRC and the United States should be behaving like confident great powers, not like insecure and tragically flawed players in an ancient Greek drama.
Gender, Party, and the Press: Media Coverage of the Second “Year of the Woman”

Julia Press, Tufts University

Abstract

The 2018 election cycle was marked by many as the second “Year of the Woman,” with issues of gender on the political forefront. This study seeks to examine whether media coverage of 2018 candidates was affected by this gendered climate, and whether coverage patterns varied by candidate party. My content analysis of 745 articles suggests that in-depth media coverage of female candidates has significantly improved in almost every dimension, with qualitative and quantitative measures indicating equal coverage, and assessments of campaign resources suggesting even more positive coverage of women than their male opponents. Media coverage reflected the nature of each individual race rather than gender-based assumptions about particular candidates. However, articles centered around gender indicate an ongoing focus on the unique difficulties faced by female Republicans and Democratic women with intersectional identities. This suggests the continuing struggles female candidates confront despite “fair” media coverage, or at least the media’s focus on these supposed struggles.

Introduction

Hailed by many as a second “Year of the Woman,” the 2018 midterms saw a record number of female candidates running for office amidst a highly contested national climate of gender-based conflict.¹ In the heat of the campaign cycle, in late September 2018, Supreme Court Justice nominee Brett Kavanaugh faced accusations of sexual assault. These allegations and the contentious confirmation hearing that followed built on the momentum of the #MeToo movement, in which women across the country raised gender-based violence as a key social issue and called out perpetrators of sexual harassment.

For many scholars following the election, it was no coincidence that the gendered debates playing out in the political arena sparked an increase in the number of female candidates. This pattern mirrored that of the previous “Year of the Woman” in 1992, when a then-record number of female candidates ran for office in light of Anita Hill’s sexual misconduct allegations against Clarence Thomas, and his subsequent confirmation to the Supreme Court.² In 1992, not only did more female candidates run, but these candidates also received a groundbreaking amount of press coverage and won at higher than usual rates, a phenomenon that scholars called the “Anita Hill effect on women’s political chances.”³ The 2018 elections, too, brought an unprecedented number of female candidates, with a historic 257 women running for Congress.⁴ Moreover, the partisan breakdown of the 2018 female candidates suggests a similar correlation between Kavanaugh and
female motivation to run, as it was mainly Democratic women who entered the races. Of the 257 women who won their Congressional primaries in 2018, only 59 (22.96 percent) were Republican.5

It is clear that the national political climate encouraged female candidates, particularly Democrats, to run for office. Would that climate similarly influence media coverage of these candidates? This is a vital concern given that the media plays an outsized role in communicating candidates’ messages to the public, thus shaping voters’ perceptions of these candidates. Coverage of female candidates has historically been less substantive and less representative than that of men.6 Yet given the #MeToo movement, Kavanaugh’s confirmation, and the highly anti-Trumpist rhetoric of many 2018 candidates, perhaps this year would be different.

Moreover, I wondered whether there might be coverage variation by party. Gender gaps in representation have improved since the early days of U.S. politics, but this has largely occurred along party lines, with the Republican party remaining overwhelmingly male-dominated.7 Would this historic gap and Trump’s history of sexism impact media coverage of the women running to represent the Republican party? By examining a female Democrat vs. female Republican race, this study seeks to examine intra-gender, inter-party media coverage. Moreover, by looking at female Democrat vs. male Republican and male Democrat vs. female Republican races, it seeks to assess the degree to which existing stereotyped patterns of coverage continue to persist across parties and genders. Through a case study rooted in the historical context of the second “Year of the Woman,” this study seeks to answer the question, “How does media coverage of candidates vary based on gender and party?” Although existing research notes the variance in stereotyped perceptions of Democratic and Republican female candidates, to date no studies have examined whether these gaps appear in news media coverage.8

My findings suggest that in-depth coverage of female candidates has significantly improved in almost every dimension, with qualitative and quantitative measures indicating equal coverage, and assessments of campaign resources suggesting even more positive coverage than their male opponents. My findings indicate that media coverage reflects the nature of each individual race rather than gender-based assumptions about particular candidates. I also found that coverage of candidates’ personality traits no longer reflects their gender, as female candidates received more coverage of male than female traits, and more coverage of male traits than their male opponents. Female candidates are no longer dismissed as novelties, but are considered serious competitors in media. Lastly, I found that articles primarily focused on gender typically discussed the difficulties female Democrats face due to intersectional identities, or the difficulties female Republicans face due to their conflicting gender and party cues. This indicates the ongoing struggles female candidates confront despite “fair” media coverage, or at least the media’s ongoing focus on these supposed struggles.

My findings indicate that the gendered coverage gaps identified in previous literature may be waning, but my qualitative results stress the salience of hegemonic gender norms and media profit motives in dictating the coverage that female candidates continue to receive.

Theorizing Gender

A symbolic interactionist approach views gender as a routine accomplishment, produced and reproduced through everyday actions.9 As West and Zimmerman wrote, it is “doing gender” that creates the notion of binary differences between men and women, rather than some intrinsic, biological quality.10 Gender differences are formed through social interaction, and then reinforced to establish “the ‘essentialness’ of gender” as a category by suggesting that differences between men and women are “fundamental and enduring dispositions.”11

This process of ‘doing gender’ holds people accountable to the traditional traits and behaviors associated with sex categories. These common-sense expectations undergird gender hegemony. Gramsci first described hegemony as the normative ideal imposed by a dominant group on society, one that becomes generally accepted as the norm12. In gender, “hegemonic masculinity” is the currently accepted strategy used to legitimize and uphold patriarchy, or “the overall subordination of women and dominance of men.”13 Hegemonic femininity, then, is the accepted gender behavior by which women operate within and become complicit in upholding patriarchy.14 Non-hegemonic behaviors and qualities are characterized as “deviant and stigmatized,” coercing both men and women to
operate in hegemonic ways. Hegemonic masculinity and femininity are constantly evolving, allowing them to rework deviant behaviors into subordinated, marginalized, and complicit roles within the hegemonic gender framework. In this way, hegemonic masculinity and femininity work in conjunction with “doing gender,” the everyday interactions that shape gender norms, to reinforce patriarchy. Hegemonic gender expectations result in stereotyped notions of the interests, competencies, traits, and beliefs of men and women.

Political Gender Stereotypes

Beyond shaping the ways people operate on a daily basis, the stereotypes that emerge from hegemonic notions of gender can impact the public’s perception of political candidates. People rely on gender stereotypes to form judgments and make assumptions about others, particularly when they lack much other information. These gender stereotypes can be particularly salient in politics, when people are asked to make judgments about candidates and often do so based on little information.

The gendered notions on which people rely as cues can be broken down into three categories: trait stereotypes, issue stereotypes, and ideology stereotypes. Research has found that certain traits are considered more common among men and others among women, and these “male” traits are perceived as more important for political candidates. Voters rely on stereotyped notions of female candidates’ traits more consistently than those of male candidates. Similarly, people often hold stereotypes about the issue areas men and women are more interested in and capable of handling, with male issues deemed more important for political candidates at all levels of office. According to some researchers, the salience of female issue stereotypes can present an opportunity for female candidates. By focusing their campaigns on male issues, female candidates can present themselves as having a broader range of expertise, because voters will assume a proficiency on female issues whether or not they are explicitly emphasized. Additionally, many studies have found that respondents assume candidates’ ideological positions based on their gender, with women of both parties presumed to be more liberal than their male counterparts.

These stereotyped assumptions can differential-ly impact female candidates of different parties. Voters perceive male traits as more Republican and female traits as Democratic, which, according to existing research, causes Democratic and Independent voters to support female candidates, while Republicans are less likely to support them. Stereotype usage in assessing Republican female candidates has an overwhelmingly negative effect, regardless of the party of voters. Ideological stereotypes in particular can influence voters’ perceptions of Republican female candidates. By sending conflicting cues about their ideology, with their gender indicating liberal and party indicating conservative, Republican female candidates confuse voters.

Another set of stereotypes on which voters rely in making assessments of candidates is race. Historically, Black voters have overwhelmingly supported Democratic candidates, so the ideology stereotype has developed that Black candidates are more liberal than their White counterparts. Hand in hand with these ideological stereotypes come issue stereotypes about Black candidates, whom voters perceive as better able to help poor and minority groups. Experimental studies have found that White voters view Black and Hispanic candidates as less competent but more compassionate than White candidates. In a 2011 study, Black politicians were more often described as opinionated, passionate, and vocal, and less often described as educated, powerful leaders compared to politicians in general. Respondents in the study viewed Black politicians as more competent in areas of minority issues, welfare, unemployment, and urban problems, and general “politicians” as adept at issues regarding the economy, national security, terrorism, and education. These assumptions can harm candidates of color in races where the latter group of issues forms the cornerstone of debate.

The researchers noted that both female candidates and candidates of color are typically deemed more honest than White male candidates, and they hypothesized that these stereotypes would affect electoral outcomes. Indeed, they found that when choosing between Barack Obama and John Edwards, the White
male candidate, voters were more likely to rely on honesty assessments, but when comparing Obama to Hillary Clinton, they used judgements of leadership and intelligence. The researchers suggested that, because both women and minorities are stereotyped as honest, voters turn to other trait stereotypes in attempting to distinguish between these candidates.35

Although both gender and race act as cues in voter assessments of candidates, it is impossible to discern the separate roles these identities play. Social systems of power, like gender and race, operate in conjunction with one another, complicating the experiences of people with intersectional identities. For instance, the testimony of Anita Hill, whose experience defined the first “Year of the Woman,” cannot be understood merely as an instance of violence against women. For women, rape embodies gendered oppression, but for African American men, rape has historically been leveraged as justification for lynching. In fact, Justice Thomas referred to his hearing, and thus Hill’s testimony, as a “high-tech lynching.”36 As an African American woman, then, Hill found herself at the intersection of these contrasting narratives.37 African American women face particular “controlling images,” gendered and racedracialized stereotypes that define how they are understood by society and maintain their subordination.38 Differences in the way members of subordinated racial and class groups “do gender” justify their subjugation, effectively excluding non-White women and men from gender hegemony.39 Though this study is primarily interested in examining the role of gender in political media, it cannot separate gender from candidates’ other marginalized identities: in this case, looking at Stacey Abrams as a Black woman and Kyrsten Sinema as a bisexual woman.

Gender in the Media

Voter assessments of candidates are profoundly shaped by the information voters receive through news media consumption. Through agenda-setting—signaling what issues are most important—and priming—dictating the standards on which the public should rely when making political evaluations—the media is able to shape what issues and dimensions of issues the public considers.40 Issue framing, another media effect, shapes the way the public thinks about and understands the subjects that have been put on the agenda.41 Through these processes, the media is able to dictate both what issues are thought about and how they are thought about. This power to shape conversations is crucial in elections, when much of the public relies on the media to receive information about potential candidates. Historically, journalists have overlooked female candidates in their coverage of political campaigns, with male candidates receiving more media coverage than their female opponents.42 Fridkin and Kenney found that news coverage privileges male candidates in the number of paragraphs devoted to them, the prominence of this coverage, and the degree of accuracy with which the media reflects candidates’ campaign messages.43 Some research has come to the contradictory conclusion that female candidates receive equal or even more coverage compared to their male opponents.44 These conflicting results may be explained by the limitations of case study approaches.

However, underrepresentation is only one component of media sexism, which Haraldsson and Wängnerud defined as “the (re)production of societal sexism through under- and misrepresentation of women in the media, leading to a false portrayal of society through a gendered lens.”45 In addition to silencing women’s voices, journalists both perpetuate and create new stereotypes that limit their ability to be upwardly mobile by enforcing restrictive, hegemonic notions of their gender.

The media often focuses on female traits in relation to female candidates, rather than accurately reflecting their campaign messages that focus on both male and female traits.46 Discussion of gender, appearance, children, and marital status has also historically appeared more frequently in coverage of female candidates than that of their male counterparts.47 However, the media continues to privilege male traits overall, focusing on them more in coverage, particularly that of races featuring two male candidates.48 The media’s overall focus on male traits, particularly in relation to male candidates, could reflect a tendency to privilege male traits in evaluating candidates, as men have historically been those running for office. However, gendered trait frames on which reporters rely may be waning.49

Besides a difference in the traits attributed to women and men, there is evidence of a difference in the amount of attention given to traits, as opposed to issues, in the media. Male candidates often receive
more issue-based coverage than female candidates, which can disadvantage female candidates by depriving voters of information on their policy positions. Despite female candidates’ attempts to emphasize issue positions in their campaigns, their media coverage does not reflect this effort.

On the other hand, some studies have indicated that the quantity of trait and issue coverage is more equal among male and female candidates. However, some of these studies revealed other ways in which media coverage might be working against female candidates, such as focusing on male traits overall and describing female candidates with more female than male traits. One study, comparing coverage of global leaders, found that the press covered the traits and issue priorities of both men and women in a nuanced way, indicating that perhaps coverage gaps are unique to American journalism.

When female candidates do receive issue coverage, this coverage often reflects stereotypical female issues. According to several scholars, despite an increased focus on male issues and traits in their campaign advertisements, female candidates continue to receive less male issue and trait media coverage than their male counterparts. The media fails to accurately reflect the messages on which female candidates focus their campaigns, but does, however, faithfully portray male candidates’ messages. Overall, reporters focus on male issues in coverage of campaigns, which can harm female candidates who benefit electorally when their races emphasize female issues.

When female candidates’ issue positions are covered, the media directly quotes the women less frequently and includes less evidence supporting their positions. Fowler and Lawless found that although female candidates received slightly more coverage of their issue positions than men, coverage of male candidates was almost twice as likely to mention concrete actions, while women were more likely to receive personal trait coverage.

Another area of coverage disparity is horse race coverage, focused on the competition and candidate viability assessments rather than candidates’ policy positions. Female candidates receive more horse race coverage than their male counterparts. Scholars have found that this horse race coverage tends to disadvantage female candidates by assessing their viability as a candidate in a negative light. Some research has tempered this notion of female candidates’ harmful horse race coverage, indicating improvements in races for certain levels of office or in more recent years’ general elections. Horse race viability assessments send explicit messages to voters telling them how seriously to regard female candidates’ chances of election, thus they provide a crucial indicator of media coverage quality. Evidence is inconclusive on whether these differences in media coverage of male and female candidates have the potential to negatively impact female candidates electorally. However, in Kahn’s 1994 experiment, Senate candidates who received prototypical male media coverage were more likely to win their race. Moreover, media sexism may be negatively impacting women’s willingness to run for office in the first place, as one recent study found.

Building on the foundation of previous research, this study seeks to fill the gaps in stereotypical media coverage of candidates based on party and gender. Although there is no consensus on whether potential differences in coverage harm or help female Republican candidates, it is clear that female candidates are subjected to different stereotyped assumptions based on their party and these assumptions may be reflected in media coverage. This study seeks to elaborate on other scholars’ work and apply it to the unique political climate of the 2018 midterm elections, while also fleshing out the role that partisan differences may play in gendered media coverage.

Methods

This study employs a media content analysis of three highly contested races, which represent the two levels of office that received the most significant media coverage during the midterms: senatorial and gubernatorial races. Incumbency is a powerful factor in shaping election outcomes, so I chose to only sample races that had no incumbents, eliminating that variable so as to more clearly examine the impact of gender and party on these races. Each race had at least one female candidate, and the races I selected represent each of the possible gender-party matchups: one with a female Democrat vs. a male Republican, one with a female Republican vs. a male Democrat, and one with two female competitors, a female Democrat vs. a female Republican.

I selected the Georgia gubernatorial race
(Democratic woman Stacey Abrams vs. Republican man Brian Kemp), the Tennessee Senate race (Republican woman Marsha Blackburn vs. Democratic man Phil Bredesen), and the Arizona Senate race (Republican woman Martha McSally vs. Democratic woman Kyrsten Sinema). Although these races do not represent every geographical region of the country, I prioritized competitiveness, incumbency, and candidate gender/party breakdowns over geographical diversity, as the former qualities are more relevant to this study. Moreover, these three “red” states all voted for Trump in the 2016 presidential race, presenting an interesting case study in a year when many candidates and voters viewed the midterm races as a referendum on Trump’s presidency.68

I selected publications to sample based on circulation statistics and accessibility through the Nexis Uni database. For mainstream national publications, I selected The New York Times, The Washington Post, and the Associated Press, all of which offered significant coverage of these races. The former two publications are among the most widely circulated publications in the U.S., and content from the Associated Press is syndicated in newspapers across the country.69 I also selected the newspaper with the largest circulation in each state that was accessible via the Nexis Uni database: The Atlanta Journal-Constitution, the Tennessean, and the Arizona Capitol Times. Additionally, I looked at television news transcripts from Fox News, CNN, and MSNBC, to assess the coverage of partisan television sources.70

I pulled all articles in the time frame from September 1, 2018 through election day, November 6, 2018, to capture the traditional campaigning period. For each news source in which a candidate appeared in over 30 articles during this time frame, I performed a simple random sample of the articles to select 30 for my analysis. Before conducting my simple random sample, I removed any duplicate articles from my sampling frame. I omitted stories that were not structured as articles, but rather as field notes from campaign rallies, nor did I include stories where a candidate was referenced fewer than three times, except when the article's primary focus was gender.

My sample resulted in a total of 745 articles coded. Of these stories, 458 focused on female candidates and 287 on male candidates. Broken down by candidate, Abrams was featured in 190 articles, Kemp in 179, Blackburn in 119, Bredesen 108, McSally 72, and Sinema 77. When looking at the sources themselves, 401 articles came from nonpartisan news sources, compared to 228 left-aligned sources (including CNN) and 116 right.

I developed measures intended to capture stereotyped patterns identified in previous literature regarding trait, issue, ideology, and horse race coverage, in addition to quantitative measures of an article’s focus on the candidate. For all decision rules and details, see the full codebook in the Appendix. In addition to my quantitative coding, I took note of any particularly gendered discussions or terms for later qualitative analysis.

**Equal Quantity, Representative Issues, and Reversed Trait Trends**

Coverage of female candidates has significantly improved in almost every qualitative and quantitative dimension, compared to findings in many prior studies. I found no statistically significant distinctions in the percentage of an article that covered each candidate or the percentage of article headlines that mentioned the candidate.71 Overall, my data show that female candidates receive significantly more issue-based coverage than male candidates. On average, female candidates’ positions on issues came up in 1.56 paragraphs per article, compared to only 0.94 paragraphs on male candidates’ issue stances (p=0.0001). However, when these results were broken down by candidate, the significant gaps disappeared. This disappearance suggests that there are no broader gender-based differences in coverage advantaging female candidates, but also indicates the absence of expected patterns of coverage quantity working against women.

My findings suggest that gendered issue coverage is correlated more with the nature of each race than the gender or party of each candidate. The only race in which trends in opposing candidates’ gendered issue coverage did not closely align was the Abrams/Kemp race, which is explained by Abrams’ focus on minority rights, coded as an issue stance, and Kemp’s suppression of minority voters, which could not be coded as an issue stance. The candidates in the two other races, however, had strikingly similar patterns of gendered issue coverage. Both Tennessee candidates mentioned more female issues than male, and both
Arizona more male than female.

This difference in gendered issue coverage can be explained by local debates. The Arizona Senate race was heavily focused on immigration and foreign policy, particularly as McSally is a former Air Force pilot and Sinema had previously spoken out against the wars in Iraq and Afghanistan. Coverage of the Tennessee Senate race, on the other hand, focused more on abortion and Kavanaugh's confirmation, driving female issue coverage. These findings suggest that media coverage of candidates has improved, with a focus on issues representative to the race rather than driven by assumptions stemming from a candidate's gender.

My findings corroborated previous research that indicated an overall focus on male traits in the media. This may be attributed to the traditional reliance on male traits in evaluating candidates, since political candidates have historically been male. However, although both male and female candidates were more often described by male than female traits, I found that female candidates received significantly more male trait coverage than male candidates did (p=0.045). Although these gaps did not remain consistent when looking at the individual candidate level rather than comparing by gender overall, the equal or greater appearance of male traits in relation to female candidates indicates an important shift in gendered coverage patterns.

The finding that female candidates as a group received more male than female trait coverage and more male trait coverage than their male opponents indicates a significant development in the representation of female candidates in the media, with previous gendered trait trends completely reversed. Perhaps this indicates a tendency among reporters to perceive male traits more strongly in female candidates, as they undermine hegemonic expectations of how a woman should act. However, in the domain of politics, the perception of male traits as more important to make a good leader suggests that this heightened focus on their male traits would benefit female candidates.

Stereotypical ideological assessments were also absent in these articles. Additionally, no notable gendered gaps appeared when comparing the primary focus of each candidate's coverage. However, it is interesting to note that only the candidates in the Arizona race, McSally and Sinema, had the greatest percentage of their coverage focused on issues. This perhaps indicates that when gender differences between the candidates are eliminated, reporters are more likely to focus on the issues of the campaign.

### Perceptions of Candidate Viability

Another dimension in which coverage of female candidates has seemingly improved is the portrayal of a candidate's qualifications, campaign successes, and chances of being elected. Unlike previous studies, which found that female candidates are presented as novelties rather than serious competitors, my data suggest otherwise. On average, female candidates were assessed as having better chances of winning than male candidates. This finding is statistically significant, indi-
cating that female candidates were not only deemed as competitive, but even more so than their male oppo-
nents.

Table 2: Average Viability Assessment by Candidate Gender

<table>
<thead>
<tr>
<th>GENDER</th>
<th>Average viability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>4.22 ± 0.11</td>
</tr>
<tr>
<td>Male</td>
<td>3.87 ± 0.13</td>
</tr>
<tr>
<td>Difference</td>
<td>0.35***</td>
</tr>
</tbody>
</table>

However, when broken down by candidate, this trend disappears. Kemp was deemed as having a significantly better chance at victory than Abrams (p=0.0001), and Blackburn better than Bredesen (p < 0.0001). There were no statistically significant differences in the Arizona Senate race. Although there no longer appears a female advantage when examining individual case studies, these findings also indicate the lack of a male viability assessment advantage, as would have been expected based on previous research.

Table 3: Average Viability Assessment by Candidate

<table>
<thead>
<tr>
<th>RACE</th>
<th>Viability assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>RACE 1</td>
<td>Abrams</td>
</tr>
<tr>
<td></td>
<td>3.85 ± 0.09</td>
</tr>
<tr>
<td>Kemp</td>
<td>4.14 ± 0.12</td>
</tr>
<tr>
<td>Difference</td>
<td>-0.29***</td>
</tr>
<tr>
<td>RACE 2</td>
<td>Blackburn</td>
</tr>
<tr>
<td></td>
<td>4.75 ± 0.22</td>
</tr>
<tr>
<td>Bredesen</td>
<td>3.76 ± 0.41</td>
</tr>
<tr>
<td>Difference</td>
<td>1.18***</td>
</tr>
<tr>
<td>RACE 3</td>
<td>McSally</td>
</tr>
<tr>
<td></td>
<td>4.02 ± 0.21</td>
</tr>
<tr>
<td>Sinema</td>
<td>4.18 ± 0.24</td>
</tr>
<tr>
<td>Difference</td>
<td>-0.15</td>
</tr>
</tbody>
</table>

Moreover, the amount of horse race coverage allotted to each candidate seemed to reflect the nature of each electoral race, rather than the candidate’s gender. Within each race, candidates received very equitable amounts of horse race coverage, both in average number of paragraphs per article and in percentage of articles overall that primarily focused on the horse race. There were no statically significant differences in horse race coverage for any of the races. This finding indicates that female candidates may no longer suffer from a lack of substantive content, with articles focusing on the horse race rather than their issue positions.

Even more noteworthy are the differences between mentions of a candidate’s positive or negative resources. The gendered breakdown of resource allocation is virtually a mirror image, with female candidates assessed as having nearly double the amount of positive resources as negative, and male candidates having nearly double the amount of negative resources as positive.

Table 4: Average Mentions of Resources by Candidate Gender

<table>
<thead>
<tr>
<th>Gender</th>
<th>Positive resources</th>
<th>Negative resources</th>
<th>Difference between positive and negative resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>0.74 ± 0.1</td>
<td>0.41 ± 0.06</td>
<td>0.33***</td>
</tr>
<tr>
<td>Male</td>
<td>0.45 ± 0.09</td>
<td>0.74 ± 0.08</td>
<td>-0.29***</td>
</tr>
<tr>
<td>Difference</td>
<td>0.29***</td>
<td>-0.33***</td>
<td>n/a</td>
</tr>
</tbody>
</table>

These findings hold true on the level of individual races, as well. Nearly all the female candidates I studied received significantly more positive mentions than negative, while male candidates were more likely to be associated with negative resources. The only candidate for whom this pattern went in the opposite direction was Kyrsten Sinema, whose media coverage was highly focused on her controversial comments as a Democrat running in a red state. I would argue that this is the exception rather than the rule, as the two other races both showed statistically significant differences between the mentions of positive and negative resources for each candidate, falling along gendered lines.
These results are particularly notable, as female campaigns were not only deemed as successful but even more so as compared to their male counterparts.

Gender-Based Coverage

In addition to performing quantitative data analysis, I coded all articles that were primarily focused on the candidate’s gender, even those that had fewer than three mentions, in an attempt to qualitatively gain a clearer picture of the media’s representation of gender in the 2018 midterm elections. This made up 21 of the 745 articles in my sample, and many of the 21 did not provide significant coverage, instead briefly mentioning that a candidate sought to be the first woman in her office. Those articles that did provide in-depth gender-based coverage revealed interesting trends, with a focus on female Democrats navigating intersectional identities and female Republicans navigating conflicting gender and party cues.

When broken down by candidate, I found that all of the stories except one were focused on female candidates. A single Associated Press story, “Bredesen seeks votes from women in Tennessee Senate race,” focused on Phil Bredesen’s attempts to capture female votes by founding a “campaign-within-a-campaign” to “highlight issues important to women, such as health care costs for their families.” Though the story linked Bredesen, a male candidate, to stereotypically female issues, seemingly undermining gender norms, it only did so by reinforcing these very same norms in calling health care and family issues “important to women.” Moreover, Bredesen himself coopted these female issues, creating a narrative in which he could capitalize on them, while still portraying them as women’s issues. In this way, Bredesen remained complicit in upholding notions of hegemonic masculinity and benefited from the very thing that his campaign seemingly undermined: the idea that certain issues are meant for women to champion.

The candidate who received the most gender-based coverage was Marsha Blackburn, Bredesen’s opponent, with seven stories focusing on her gender in relation to her candidacy. Most of Blackburn’s coverage highlighted the conflicting cues she sends as a conservative female candidate. One Associated Press story, “Tennessee women slow to back GOP’s Blackburn in Senate race,” noted her difficult positioning amidst a female population that is becoming increasingly liberal, but it also described instances in which her campaign sought “to play the gender card” by drawing on her “attempts to break the glass ceiling.” It even quoted Blackburn herself, who said “Republican women are never going to get the attention that Democratic women are going to get...I fully understand that Republican women do not fit the narrative that many in the media would like to construct.”

A New York Times story described Blackburn as “invoking her gender without being prompted,” but subsequently retreating “from what could be seen as making any claim of misogyny” by claiming that “skepticism” about her candidacy stems from unfamiliarity, not sexism. The theme of walking a fine line between her conservative and female sides appeared across the nonpartisan stories I coded. Another Times story noted that Blackburn “chose to calibrate her response” to the allegations against Brett Kavanaugh, carefully denouncing the Democrats’ accusations while noting the importance of listening to all women who speak out about such allegations. An article from The Washington Post, “For women on the 2018 campaign trail, ‘sexism’ is no longer a forbidden word,” suggested that Blackburn has been unique among Republican candidates in how she “has leaned into gender confrontations” with her opponent, considering Trump’s popularity with the party. Yet the story also noted the increasing group of voters for whom candidate gender matters.
Since the bulk of Blackburn's gender-based coverage focused on her as a female Republican, I wondered whether this would also be the case for McSally, as a Republican woman running against another female candidate. Although McSally and Sinema received the same amount of gender-based coverage (five articles each), I sought to understand whether the content of their coverage differed and whether that of McSally resembled that of Blackburn. Unfortunately, coverage of McSally was not substantive—all of the articles that focused on her gender consisted only of brief mentions, often noting that both she and Sinema were running to become Arizona’s first female Senator. Out of the case studies in my sample, this is the race that received the least amount of media coverage overall. Thus I can only posit that Blackburn’s pattern of gender-based coverage focusing on her conflicting gender and party cues would hold true for other female Republican candidates.

The one article that provided in-depth gender-based coverage of the Arizona race focused on Sinema, not McSally. The story, entitled “Women, minority candidates emerge in Democratic slate,” was published in the Arizona Capitol Times. The piece quoted a Republican political consultant, who described the importance of candidates appealing to a broad range of voters across the ideological spectrum. He cited Sinema as an ideal example of someone “is not relying on the female or LGBT candidate equation to mean victory.”78 Instead, she and other strategic “diversity” candidates are moving beyond “identity politics” to connect with voters. Interestingly, this article suggested a potential harm that could come from too great a focus on one’s marginalized identities, stating that “diversity alone is not enough” to win. However, unlike the previous article that suggested Sinema’s gender and sexuality were easily contained under the banner of “diversity” candidate, this story explored the nuanced struggles that Abrams must face as a female Democratic candidate of color when non-Black women fail to support her.

Another article on Abrams, “Women who become governor will have to manage expectations,” which appeared in the Associated Press, presumably focused on gender, but in fact just highlighted the obligatory give-and-take between all governors and their state legislatures. Its only relation to gender was covering female governors like Abrams, whom it described as championing issues such as “abortion, marriage equality and voting access.”79 This article, unlike the others, seemingly presented being a woman as, in and of itself, a candidate’s defining quality that would influence the issues she champions and her negotiations with others once in office.

My sample of articles that focused on gender in-depth indicate the continued attention paid to female Republicans’ conflicting signals and female Democrats’ intersectional identities. This finding suggests that even if media coverage is not biased against them, these candidates continue to struggle with getting their campaign messages across—or at least that journalists present such a conflict.

**Gendered Discussions**

Even in articles less directly focused on gender itself, certain discussions indicated patterns of female can-
Gender, Party, and the Press

didate coverage that would likely not appear for male candidates. For instance, an MSNBC report from September 2018 primarily focused on McSally’s need to improve her “likeability,” a quality never discussed in relation to any of the male candidates I studied and reflective of past research on the undue focus on female candidates’ personalities. Similarly, although there was vast improvement in the dimension of appearance-based coverage, the one article that featured descriptions of candidates’ outfits was an Associated Press story on the Arizona Senate race. This piece, “Campaigns make final push to galvanize voters for midterms,” opened by noting that McSally “wore a maroon-and-gold Arizona State University T-shirt and jeans,” while Sinema “wore a canary-yellow dress and 4-inch platform shoes with cactus applique.” Though revealing nothing about the content of the candidates’ campaigns or the issues for which they stood, the article sought to position the candidates as starkly different through focusing on their appearances. This sort of appearance-based coverage was seen in none of the coverage of other campaigns. I suspect that it remains here only because both candidates were female and the juxtaposition of their outfits seemed to the reporter an apt way to illustrate their different personalities. This technique perpetuates the historic, aforementioned pattern of focusing on female candidates’ superficial qualities over substantive issue coverage, and establishes inequitable expectations for candidates of different genders. By assuming that their outfits speak to their characters, the article draws on hegemonic expectations of femininity and compels the female candidates to “do gender” by carefully considering their outfit choices, as they will be interpreted by the press, while men face no such scrutiny.

In certain cases, gendered discussions were driven by the candidates themselves. McSally sought to harness gender to belittle her opponent by asserting, “while we were in harm’s way, she was protesting our troops in a tutu.” McSally used the image of a tutu to evoke Sinema’s femininity and juvenility, suggesting weakness and a lack of leadership potential. Juxtaposed against McSally’s own description of herself as a former Air Force combat pilot, this image of Sinema appears particularly disparaging.

In other instances, female Republican candidates used their gender to justify policy decisions that a woman might not be expected to take. When describing the need for heightened border security, Blackburn described hearing from mothers about their children’s safety, rather than taking a stereotypically masculine stance focused on crime and defense. She cited “the security mom in every one of us” as the driving force behind immigration crackdowns, seeking to reconcile the apparent conflict between her issue stance and her gender. Similarly, McSally noted the importance of supporting “Angel Families,” relatives of those killed by undocumented immigrant crime, using family to strengthen her position on immigration restrictions. Despite a national conversation linking strict immigration policy to the separation of children from their parents at the border, these female Republicans flipped the narrative to focus on families as the reason to support immigration crackdowns. As two white female candidates, Blackburn and McSally are considered protectors of the “blood ties” that bind the American nation, or national family unit, together. Their privileged racial position lets them use their gender to promote ethnic nationalism and American hegemony, advancing conservative policies by drawing on their identity.

Discussion

According to my findings, the media has vastly improved its coverage of female candidates, providing readers with a more equitable and representative look at their campaigns than past research has indicated. I did not find evidence of reporters playing into hegemonic notions of gender in their coverage, nor did they disparage the legitimacy of female candidates and their campaigns. This improvement has widespread and significant repercussions for the success of female candidates in the political arena, where they have historically been dismissed, belittled, and overlooked. Taken with Kahn’s finding that Senate candidates are more likely to be elected if they receive stereotypical male media coverage, these strides toward equal coverage of candidates, regardless of gender, can level the playing field and open up more political opportunity to women.

However, my findings about articles that focused primarily on gender revealed striking trends lost in the quantitative coding data, suggesting that being a woman alone does not dictate media coverage. Rather than simply facing the challenge of entering a pre-
dominantly male field, Abrams and Sinema, bound by overlapping systems of oppression, confront additional hurdles to appealing to a broad voter base. Research on Trump’s 2016 election victory suggests that racism and sexism played a significant role in driving support among his base.87 For a female candidate of color or female bisexual candidate, then, holding multiple marginalized identities can be a disadvantage when facing the inherent biases that voters hold. My findings indicate that these candidates not only suffer electorally in this way, but may also struggle with how to present their identities to voters through the media. Discussion of Sinema’s strategy of not emphasizing her sexual identity too heavily and Abrams’ difficulty in capturing a White female electorate highlight the conflict the media presents when covering these candidates. Whether or not these identities are truly harder for the public to comprehend or accept, the media generates a narrative of conflict that could shape voters’ perceptions and influence votes.

It seems as if the media’s search for stories driven by novelty, controversy, or conflict has shaped the gender-based coverage female candidates across the aisle have received. After years of industry dominance by the “Big Three” major television news networks, the recent proliferation of media outlets, audience fragmentation, and technological advances have generated the 24-hour news cycle that exists today.88 In this highly competitive environment, news outlets have been forced to develop models to draw in audiences and generate a profit. This economic necessity led to the emergence of news “infotainment,” in which relaying information is secondary to the goal of including “drama, emotion, plot simplicity, personal morals, and character conflicts.”89 Profit motives incentivize “substituting controversy for reporting” and push even renowned news organizations towards “sensationalism.”90 Moreover, research has indicated that episodic frames, in which a story focuses on a particular narrative, are more accessible to audiences than thematic frames, which focus on broader issues.91 These imperatives can push reporters to seek out compelling, conflict-driven narratives in their political reporting.

In 2018’s “Year of the Woman,” being a female candidate did not make someone noteworthy. Instead, in an effort to tell compelling stories that draw in an audience, the media shifted its focus to other conflicting or marginalized identities these candidates held. While there was no greater degree of scandal-driven coverage for female candidates, my qualitative findings suggest that reporters chose to highlight conflict-driven stories in relation to gender. For this reason, although my quantitative measures suggested fair and representative coverage of candidates across gender and party lines, my qualitative findings on gender-focused coverage indicate that the media continues to focus on the particular struggles female candidates face.

For female Republicans, too, a qualitative examination of articles focused on gender revealed the ongoing conversations that are occurring about their conflicting messaging and their struggles to be easily understood by or appeal to a wide base of voters. Gender remains a salient cue offering stereotyped notions of hegemonic femininity that are not reflected in these candidates’ conservative party. Female Republicans fail to fit neatly into gendered norms, and by undermining hegemonic notions of what it is to be a female political candidate, they confuse voters and reporters alike. Ideological stereotypes reveal that women are typically assumed to be liberal, yet Republican women champion conservative policies and sometimes present a tough, “masculine” façade. McSally asserted a strong pro-war stance and emphasized her past as an Air Force combat pilot; Blackburn opposed abortion and supported strict border security. These are not the positions of the liberal female candidates voters expect to find. Even in three “red” states that voted Republican in the last presidential election, media coverage reveals the confusion that these candidates continue to elicit. Voters and reporters struggle to understand female Republican candidates because they so clearly undermine the hegemonic expectations of their gender in this way.

Additionally, one quote in particular indicated that these female Republican candidates may be aware of the public’s inability to understand their norm-breaking positions. Blackburn’s statement that “Republican women do not fit the narrative that many in the media would like to construct”92 demonstrated her acknowledgment of the role conflict she faces. The limiting conceptualization of the gender binary can often result in role conflict, by which “persons of a particular sex category can ‘see’ quite clearly that they are out of place and that if they were not there, their current troubles would not exist.”93 Republican women recognize that their conservatism does not adhere
to hegemonic notions of femininity and can clearly conceptualize the barriers this creates to their being understood by voters and the press. Marsha Blackburn evidently revealed her own understanding of the conflicting cues she and others in her group send to voters, and suggested that the news media would unfairly cover them because of it. In fact, all quantitative measures suggest that Republican women receive media coverage equitable to that of candidates of other parties and genders. However, qualitative evidence suggests that Republican women do face unique obstacles that remain a media focus, which could harm these candidates by detracting from the issues or putting undue emphasis on the conflicting signals they send to voters. The electoral impact that this media focus causes could provide ample material for future study.

Despite talk of 2018 as the “Year of the Woman,” this “so-called pink wave was hardly bipartisan,” with only four new female Republicans elected, compared to 38 Democrats.94 Moreover, the photograph that circulated comparing the Democratic and Republican freshman representatives showed voters the striking difference between the two parties: the Republican photo featured a sea of White men, while the Democrats appeared a patchwork of diversity.95 Although my study identified progress in the media landscape for female candidates, the political arena remains an unequal playing field, particularly for Republican women and candidates of color.

As women, and particularly people of color, have been systematically attacked by the current presidential administration, media outlets can play an important role in shaping the conversation and working to combat persistent inequalities. If and when the third “Year of the Woman” comes, we should be focusing more on women of intersectional identities, particularly women of color. By uplifting their voices in the media and treating them as real contenders rather than controversial candidates, the press has the power to open the political arena to new voices who can champion the causes that matter to others of their identities. While the most recent Congress brought in many new representatives of color, lifting the percentage of minority Congressmembers to 22 percent, this remains highly partisan. Representatives of color made up only 33 percent of Democrats elected to the House in 2018, but only two percent of Republicans.96 A bipartisan, intersectional approach to feminism, in politics and beyond, is crucial to bring about meaningful change on the structural level. Moreover, women should be uniting across party lines to oppose the subjugation they are facing by many politicians currently in power. Getting more women elected on both sides of the aisle will help promote the best interests of all women, and fair media coverage can go a long way in elevating the voices of these female candidates and increasing their chances at victory.

However, improved media coverage is only one small step toward undoing or redoing the system of gender that privileges some at the expense of others. As a system established to maintain inequality, gender must be completely upended in order for all people to enjoy full equality.97 Improvements in media coverage may contribute to a more accessible political system, but to completely remove the “patriarchal dividend” that continues to let men benefit from the subordination of others, any notions of hegemonic gender norms must be abolished.98 Because of the intersectional nature of identity, hierarchies of race, socioeconomic status, sexual orientation, and ability must also be systematically upended to achieve a society where gender does not work to perpetuate inequality. In order to fully level the playing field both in and outside the political arena, we must push beyond these boundaries and envision a world where all identities are celebrated and hierarchies are eliminated.

Appendix

Codebook

[ID] Article ID
   ID number (i.e. 0001, 0002)

Article source

[PUB] Publication name

[HEADLINE] Headline

[URL] Link to article

[CANDIDATE] Candidate name
   (1) Stacey Abrams (2) Brian Kemp (3) Marsha Blackburn (4) Phil Bredesen (5) Martha McSally (6) Kyrsten Sinema
Candidate's ideological stance, as presented by the news source. Only code when a publication explicitly writes about a candidate's ideology or the ideology of voters who support the candidate, do not infer from issue stances.

Liberal (2) Moderate/Centrist (3) Conservative (4) n/a

[HORSEQUANT] Horse race coverage
Number of paragraphs that focus on the horse race, or coverage that emphasizes the competition, polling, or likelihood of victory

[VIABILITY] Horserace assessment of candidate
(0) n/a (1) sure loser, (2) behind but gaining ground, (3) slightly behind, (4) competitive, (5) slightly ahead, (6) ahead but losing ground, (7) likely winner

[POSRES] Positive resources
Number of mentions of the candidate's positive resources (i.e. successful fundraising, support from leaders or public figures, party advantages). Personality traits should be coded under PERSONAL, not POSRES. Qualifications that prepare the candidate should be coded under QUALS, not POSRES.

[NEGRES] Negative resources
Number of mentions of the candidate's negative resources (i.e. unsuccessful fundraising, lack of support from leaders or public figures, party disadvantages, controversial scandals)

[QUALS] Qualifications
Whether an article mentions a candidate's presence or lack of experience in office, relevant professional non-political experience, or education.

Qualifications mentioned (2) No qualifications mentioned

[PERSONAL] Personal frame
Number of paragraphs that mention the candidate's personal background and traits. This can include mentions of the candidate's identities (i.e. race, gender, socioeconomic status, sexual orientation), family (children, marriage), personal interests, and personality traits. Personality traits should only be counted under PERSONAL if explicitly mentioned, not implied as in the QUANTTRAITF and QUANTTRAITM categories.

[APPEARANCE] Appearance
Number of mentions of the candidate's appearance in the body of the article or headline.

Content of traits
Traits need not be mentioned explicitly, but inferred from context—for example, if one candidate “attacks”
another, code it as a male trait (aggressive). In interviews with candidates, include traits the candidates themselves present (i.e. “I’m tough on the border”), but not in cases when the article contradicts the candidate’s presentation of him/herself. Do not include traits that others present (i.e. “My opponent is weak on the border”).

[QUANTTRAITF] Female traits
Number of female traits mentioned in relation to the candidate. Female traits include: honest, compassionate, able to compromise, emotional, dependent, non-competitive, passive, weak leader, feminine.

NOTE: Only code competitiveness as a trait when explicitly mentioned in relation to a candidate’s character, not the nature of the race.

[QUANTTRAITM] Male traits
Number of male traits mentioned in relation to the candidate. Male traits include: able to handle a crisis, decisive, emotionally stable, strong leader, aggressive, independent, objective, competitive, ambitious, tough, masculine.

NOTE: Only code competitiveness as a trait when explicitly mentioned in relation to a candidate’s character, not the nature of the race.

[TONE] Overall tone of article
Assessment of whether the article generally uses language and quotations that are positive, approving, optimistic, favorable, or that which is angry, antagonistic, critical, unfavorable.

(1) Positive (2) Negative (3) Neutral (4) Mixed

[NOVELTY] Novelty frame
Whether the article focuses on the candidate or campaign as novel in some way (i.e. “first female representative”)

(1) Present (2) Absent (3) Not applicable

[QUOTES] Quotations
Number of paragraphs that include direct quotations from the candidate. Social media posts, excerpts from campaign advertisements, and statements from candidate proxies (i.e. campaign managers) count as direct quotes.

[PRIMFOC] Primary focus
(1) Candidate traits or background (2) Issues (3) Horse race (4) Campaigning (fundraising, strategy, support from party leaders, debate coverage, etc.) (5) Gender (6) Scandal/controversy (7) Other

Endnotes
3. Witt, Paget and Matthews, Running as a Woman.


17. For the purposes of this study, I operate within a gender binary, as all of the candidates in my sample are cisgender.


21. For an overview of which traits and issues are typically considered “male” and “female,” based on previous research, see the codebook in the Appendix.


29. Koch, “Do Citizens Apply Gender Stereotypes to Infer Candidates’ Ideological Orientations?”

30. Because this study’s sample includes only one candidate of color, a Black woman, my literature review of political racial stereotypes focuses on the differences between Black and White candidates.


32. McDermott, “Race and Gender Cues in Low-Information Elections.”


34. Schneider and Bos, “An Exploration of the Content of Stereotypes of Black Politicians.”

35. Schneider and Bos, “An Exploration of the Content of Stereotypes of Black Politicians.”


39. Schippers, “Recovering the feminine other: masculinity, femininity, and gender hegemony.”


57. Kahn, *The Political Consequences of Being a Woman*; Kittilson and Fridkin, “Gender, Candidate Portrayals and Election Campaigns.”

58. Kahn, *The Political Consequences of Being a Woman*.


65. Kahn, “Does Gender Make a Difference?”


70. In my complete study, I also examined how the identities of content creators impacted media coverage of candidates, by noting the partisanship of each outlet and the gender of each reporter. For the purposes of this article, I will only discuss variance in the gender/party of candidates, rather than also including my results on the gender/party differences amongst reporters.

71. Since my sample excluded articles in which a candidate was mentioned fewer than three times, my findings can only speak to “in-depth” coverage. It is possible that coverage gaps may appear when looking at articles with fewer candidate mentions.

72. Female candidates' female-male trait t-test had a p-value < 0.0001, and that of male candidates had a p-value = 0.0097.


84. Laura Ingraham, “Focus Group Discusses Midterm Election and Immigration; Focus Group Discusses U.S. Economy’s Possible Effect on Midterms; How President Trump’s Rhetoric Could Affect The Voter’s Mind; Cong. Martha McSally Have Laid Out Her Platform Against Her Democrat Opponent Krysten Sinema; One of Justice Brett Kavanaugh’s Accusers Recants Allegations,” Ingraham Angle, Fox News, November 2, 2018.


86. Kahn, “Does Gender Make a Difference?”


90. Bennett, News, 23.


92. Kruesi, “Tennessee women slow to back GOP’s Blackburn in Senate Race.”

93. West and Zimmerman, “Doing Gender,” 140.


Bibliography

Finneman, Teri. Press Portrayals of Women Politicians,


gress-diversity-democrats-republicans-photo.


Lühiste, Maarja and Susan Banducci. “Invisible Women? Comparing Candidates’ News Coverage...
in Europe.” *Politics & Gender* 12, issue 2 (June 2016): 223-253. https://doi.org/10.1017/S1743923X16000106.


Schippers, Mimi. “Recovering the feminine other:


Chapter 1: Background

1. Predictive Risk Models in Health and Human Services

The adoption of new machine learning techniques has promised to revolutionize a diverse set of tasks, from healthcare to finance to the distribution of welfare. Advocates of the adoption of machine learning techniques in different industries point to its ability to offer accurate results while reducing human bias in decision-making.¹ The original thought is reasonable: if a machine were given the ability to make a decision based on a set of training data, human bias would be eliminated in the decision-making process. In practice, however, human biases can still enter the algorithm at many different points, including in the design and training stages and in the human-computer interaction involved with the use of the program.

The machine learning model is created by first taking in a set of training data, finding statistical relationships between the training data and a chosen target variable, and then using these relationships to make predictions of the likely outcome of future, unknown cases. The basis of machine learning is the process of creating a model that “learns” from previously known data and predicts results for future data based on statistical relationships. Machine learning models can assist in any situation in which a human must make a decision about a future outcome.

Given the promise of improved decision-making through the use of large amounts of available data,
The use of machine learning is becoming more popular in health and human services (HHS). In HHS, a machine learning model that is used to generate a risk score for the occurrence of an adverse event is called a Predictive Risk Model (PRM).2 In many ways, HHS is a good candidate for the early adoption of these systems. Many healthcare centers have access to vast amounts of data but do not have a way to use all of the data to make the best possible choice about a patient’s health. Moreover, the decisions that healthcare professionals must make have potentially fatal consequences, and any improvement in their ability to make these decisions that rely on their prediction of the occurrence of an adverse event is valuable. In New Zealand and Australia, machine learning models are already in use to assist in predicting the likelihood of the occurrence of a disease.3 Accurate information about the risk of certain diseases can help national health organizations better allocate their limited resources. At Kaiser Permanente, a healthcare company based in the United States, and at the National Health Service in the United Kingdom, these models are used to determine an individual’s risk of future emergency hospitalization.4 In each case, the PRM promises to improve the decision making skills of a human by allowing the human to analyze more data in a timely manner than would be possible without the PRM.

Despite the promise of PRMs, institutions must consider the many ethical implications tied to the adoption of PRMs, including the repercussions of mistakes. Organizations must consider the importance of high-quality professional training required for a proper understanding of the PRM, the possibility for a PRM to disparately impact different racial and socioeconomic groups, and the difficulties of obtaining meaningful consent, among other concerns. Beyond the accuracy of the PRM, a central ethical consideration involved with its adoption in healthcare involves the user’s understanding of the limits and uses of the risk score generated by the PRMs. Unrealistic user expectations of the risk scores or complete aversion to the risk assessment system might both lead to fluctuations in accuracy and consistency of the system.

In this paper, I will conduct a case study on the Allegheny Family Screening Tool (AFST), a risk assessment tool used in child protection services in Allegheny County, Pennsylvania. In the first chapter, I

2. Allegheny County DHS

The Allegheny County Department of Human Services (DHS) houses the Office of Children, Youth, and Families (CYF). One of the roles of the CYF is to operate a call center for their local child neglect and abuse hotline. Call screeners in the center receive a call from an individual, either from the community or a mandated reporter, about potential maltreatment occurring in a family. The call screener must then search their internal database for more information about the family. After examining the available information, the call screener assigns a safety rating to the family before they come to a decision about whether to screen the family in for a formal investigation to determine if maltreatment has occurred and if there is potential for future harm, or to screen the family out without conducting any further evaluation or assessment.5 A decision to screen the family in will open an investigation in which child welfare investigators will visit the home. The investigation ultimately leads to either the opening of a new child welfare case or no further action aside from the provision of additional information for available resources, as shown in the workflow in Figure 1.6
In 1999, the Allegheny County DHS created its own data center to keep public records. Since then, it has collected over one billion electronic records from different public services. These records include historical and cross-sector administrative data over twenty years from child protective services, mental health services, drug and alcohol services, homeless services, and many other related public services. Call screeners have access to these records when they search the internal database for information about a family.

The availability of this large amount of data leads to a potential issue; call screeners cannot efficiently access and review all available records in a meaningful and timely manner. As a result, some screeners may miss important information or weight unimportant information higher based on the small sample of data accessible in the available time. Consider, for example, a variable such as the recent involvement of a parent in a criminal case. This variable might be weighted heavily by one call screener but overlooked by another. The nature of this issue suggests that the power of the data is not being effectively harnessed.

With these issues in mind, the DHS commissioned a team of researchers from New Zealand and California to create a PRM to help analyze the vast amount of available data to help call screeners make better choices. Economist Rhema Vaithianathan of the Auckland University of Technology partnered with Emily Putnam-Hornstein, the director of the Children’s Data Network at the University of Southern California, to implement the Allegheny Family Screening Tool (AFST) which uses machine learning techniques to predict the likelihood that a child will experience abuse or neglect in the future.

The DHS began use of the AFST in August 2016 to assist the decisions made by child welfare call screeners in the call center for the child neglect and abuse hotline. The purpose of the AFST is not to replace the call screener, but rather to augment the screener’s ability to make the most informed decision by analyzing the data and offering a summarized risk score of findings. As such, the AFST is a tool that must be used in conjunction with the human process to assist the human screener rather than replace the human. In an attempt to maintain human agency in the call screening process, the creators of the tool placed the tool at the end of the call screening workflow, so that the human is able to come to an independent judgement prior to running the AFST.

The overall case decision workflow remained mostly the same, with an additional step in the call screening phase, in which the call screener uses the
AFST to calculate a risk score for the family. After call screeners receive a call from the community, they still gather information and perform their normal due diligence on the case to come to an independent judgement. Only after the call screeners assign their own safety rating to the family do they run the AFST to see the risk score calculated based on the predictive model.

The resulting AFST score, a score between 1 and 20 where 1 is the lowest risk and 20 is the highest risk, can be categorized into three distinct outcomes: (1) high-risk protocol, (2) low-risk protocol, and (3) other. A score greater than 17 is categorized as a high-risk protocol, which means that the case is designated for a mandatory screen-in unless overridden by the call screener. A score lower than 11 is considered a low-risk protocol which results in a recommendation to screen the case out with the decision left to the call screener. Any score between these two ends leaves full discretion to the call screener without further recommendations.

Figure 2 captures the updated workflow for the call screener, with the placement of the additional step of calculating the AFST score highlighted in red. The creators of the tool explicitly place the AFST after a call screener has come up with her own independent risk score so that the AFST does not unduly influence the call screener’s own judgement. The idea behind this design choice is to retain the call screener’s ability to independently come up with their own judgments.

3. The AFST Risk Score

The AFST is a predictive tool that assesses a child’s risk of needing services. The score is a number between 1 and 20 where 1 is the lowest risk and 20 is the highest risk. The score for a family is determined by the maximum risk score given across all of the children in the family. For instance, if there are four children, each with a unique score, the ultimate score shown to call screeners for the family will be the maximum score among the four children. For each family, a single risk score is shown. This can be seen in Figure 3 in the form of a gradient figure that includes a colored scale where green indicates low risk and red indicates high risk.

Figure 3 depicts the screen that the call screeners interact with. After assigning their own independent risk and safety ratings, call screeners click the blue “Calculate Screening Score” button in the center of the figure. Once the button has been clicked, the program renders a blue dot with a score that is situated between the two ends of the thermometer gauge, with green indicating lowest risk and red indicating highest risk. Below the thermometer itself, there is additional information indicating the call screener who last ran the tool and a timestamp indicating when it was last run. In addition, there is an indication of the versions of the underlying algorithms used, helping the call screener understand whether the tool is using the most up-to-date models in its calculation of the score.
At the bottom of the screen in Figure 3, there is a paragraph of static text that reminds call screeners that the machine acts as an information aid rather than as a substitute for independent clinical judgement. This reminder is meant to help screeners understand the role of the machine in the context of the overall decision to investigate. The text also includes a high-level description of how the AFST calculates its score, alluding to “hundreds of data elements and insights from historic referral outcomes to estimate the likelihood of this referral resulting in the need for a child’s protective removal from the home within 2 years.”

This text reminds call screeners of two important aspects of the AFST. First, the tool uses statistical models that gather data from historical records; the scores that the tool generates are derived from the underlying statistical model. Second, and relatedly, the score estimates the likelihood that a specific referral will result in direct action on the family if an investigation is carried out. Therefore, the score is not directly measuring risk of child abuse, but rather measuring the likelihood that a decision to investigate based on a specific referral will lead to action. Both aspects of the reminder raise an important question about what the AFST is actually measuring and what the score means.

The AFST is a tool that constitutes a risk score. Although the risk score would ideally measure the likelihood of a child experiencing abuse or extreme neglect in the future, these events cannot be measured directly with the risk score because of a lack of cases of recorded abuse coupled with an ambiguity between extreme neglect and neglect. As a result, the risk score is an indirect measurement of child risk determined through an underlying proxy statistical model, the placement model.

The placement model measures the likelihood that a child experiences a safety issue that requires action on the child’s behalf, such as removing the child from their home into a safer setting within two years following a referral. In essence, the placement model is predicting the likely behavior of the investigators who conduct the investigations on the basis of referrals from the call screeners. It measures the likelihood of whether direct action is necessary based on the current referral from the investigation. By providing a prediction of the outcome of an investigation, the placement model offers a proxy for child mistreatment.

4. Evaluations of the AFST

In addition to a technical design methodology written
by the creators, the DHS has publicly released a set of three independent evaluations of the tool as part of its commitment to proceed with the adoption of the AFST in a transparent manner. The three evaluations consist of a process evaluation, an evaluation of the impact of the tool in use, and an ethical review of the tool. I will address the main findings of each of these evaluations starting with the process evaluation, then moving to the impact evaluation before finally considering the ethical review.

4.1 Allegheny County Predictive Risk Modeling Tool Implementation: Process Evaluation

Hornby Zeller Associates, a public-sector consulting firm, examined the process by which the tool was implemented. The firm analyzed the tool’s impact on the call screening experience, the practical and policy implications of its use, and the perceptions of and reactions to the tool. The study was first carried out in the summer of 2016 through interviews of call screening staff conducted prior to implementation, and multiple times after the implementation, in the fall of 2016 through the spring of 2017.13

With regard to the call screeners, the process evaluation came to two major conclusions. First, the study concluded that call screening staff report having a good understanding of the AFST and finding the tool easy to use. The majority of call screeners claim that they understand how the score works, with 100% of participants agreeing that they are “adequately prepared to use the tool.”14 The results indicate that call screeners feel extremely confident in their own ability to use the tool.

The second major conclusion of the study is that, despite apparent understanding, call screening staff vary greatly with regard to the frequency of their use of the AFST scores. The survey of call screeners two months after implementation of the AFST found that only around 40% of call screeners use the tool to inform their decisions on a consistent basis and, around 31% of call screeners reported that they rarely, if at all, use the tool.15 Despite widespread understanding of the tool's functionality, call screeners seem to differ greatly in their daily use of the tool.

4.2 Impact Evaluation of a Predictive Risk Modeling Tool for Allegheny County’s Child Welfare Office

In April 2019, Jeremy D. Goldhaber-Fiebert and Lea Prince, both faculty at Stanford University, published a report on the impact of the implementation of the AFST on call screening decisions. The study focused on the AFST’s impact on the accuracy of call screening decisions as well as on the consistency of screening decisions, both of which are stated goals of the AFST. The data was collected from August 1, 2013 through May 31, 2018 and compared the years prior to the implementation of the tool with the two years following its adoption.

In terms of accuracy, the study concludes that the implementation of the AFST increased the accuracy of families correctly screened-in for an investigation and slightly decreased the accuracy of families screened-out for an investigation. In the first case, the researchers found that a higher proportion of children screened-in for investigation result in an investigation that had further action taken. On the other hand, the researchers found a slight decrease in the proportion of children who are screened-out and had no re-referrals within 60 days.17

With regard to consistency, the study found no significant change in outcomes across call screeners. That is, the researchers report no increase or decrease in the consistency of the decisions made by call screeners between the pre-AFST and post-AFST implementation periods. However, the researchers qualify this finding with the note that there was likely an insufficient sample size to detect any minor changes.18

4.3 Ethical Analysis: Predictive Risk Models at Call Screening for Allegheny County

In addition to the impact and process evaluations, the DHS commissioned a third party to independently conduct an ethical review prior to the adoption of the tool. The researchers considered potential ethical issues that could arise from the adoption of the tool, including issues of consent for the use of information, stigmatization of families, racial disparity, professional competency and training, and the importance of ongoing monitoring.19 The purpose of this analysis was to consider a number of particular ethical issues that could arise from use of the AFST. So rather than an evaluation of the ethical standing of the tool itself, this analysis focused on particular aspects of the tool, e.g.
professional competence/training, that were essential for the tool is to operate ethically.

Ultimately, the researchers concluded that the AFST could operate ethically given that the DHS followed the suggestions offered in the ethical analysis. Moreover, the report took the even stronger claim that it would be unethical to refrain from using the AFST because of the potential for the AFST to increase the accuracy of decisions made. Given the potential for an increase in accuracy and a decrease in stigmatization that might come with the adoption of the AFST, the researchers argued that the adoption of the tool is actually morally necessary as the alternative would lead to a less accurate outcome.\textsuperscript{20} The ethicists based this conclusion on the importance of minimizing errors in child screening decisions and the unavoidability of decision-making.\textsuperscript{21}

The DHS’s decision to publish these studies demonstrates the high level of transparency with which they approached the adoption of the AFST. These studies offer valuable insight into the workings of the AFST, the workflow involved in its use, the impact of the AFST on decisions made, and also ethical considerations that the DHS are keeping in mind.

5. Controversy with AFST

Despite the DHS’s efforts at full transparency with the creation and adoption of the AFST, the tool has still been the subject of public criticism. One vocal critic, Virginia Eubanks, published a book called Automating Inequality, which offers a sociological overview of the complex issues surrounding the AFST; It provides insightful anecdotes from interviews with call screeners at the county hotline, staff at the DHS in charge of the AFST, and families that have interacted with the Allegheny County CYF.\textsuperscript{22}

Eubanks is wide-reaching in her criticisms, offering a general overview of the main issues in public discourse about the AFST including questions of discrimination, accuracy, and biases. Her book was generally well-received by the community and received numerous book awards and honorable mentions like the McGannon Center Book Prize and the Goddard Riverside Stephan Russo Book Prize for Social Justice. Moreover, her criticisms elicited a direct response from the DHS, whose response offers even greater insight into the nature of the debate.

Given the general reception of and reaction to her book, the issues raised in the book offer a valuable starting point for an ethical analysis of the AFST based on public controversy. By clarifying and building upon the issues raised by Virginia Eubanks and by mediating these with responses from the DHS, we can better understand the questions that drive the debate. My task will be to test the strength of Eubanks’ claims, arguing that Eubanks’ claims are not as convincing as they seem on further inspection.

The main charge I consider is Eubanks’ general argument that the AFST unfairly discriminates against poor families. I first offer a brief introduction of Eubanks’ argument before constructing a philosophically rigorous account of the conditions for wrongful discrimination. Then, I consider what it means for a machine like the AFST to have the ability to discriminate before evaluating whether or not the AFST fulfills the conditions of wrongful discrimination. Ultimately, I argue that the AFST does not wrongfully discriminate against poor families.

Chapter 2: Discrimination and the AFST

In this chapter, I consider Virginia Eubanks’ argument that the AFST wrongfully discriminates against certain groups of people, namely people of low socioeconomic status. Eubanks raises concerns that human biases are built into the AFST such that the resulting scores potentially discriminate against certain groups. Specifically, Eubanks suggests that the AFST targets poor families through the inclusion of public service usage in its predictive data. Part of the predictive power of the AFST comes from data about a family’s use of public services, which might initially suggest that the AFST discriminates against poor people.\textsuperscript{23} At a general level, her argument is that the AFST’s inclusion of data from the use of public services, which might initially suggest that the AFST discriminates against poor people. At a general level, her argument is that the AFST’s inclusion of data from the use of public services unfairly discriminates against families of lower socioeconomic standing.

At first glance, this argument is compelling. The AFST does include data from records of receipt of child protective services, drug and alcohol services, or homeless services offered through a public agency among other public services. Moreover, the AFST does not include any data on privately offered counterparts for these public services, such as privately hired childcare or mental health services.

Despite this initial intuition, there are import-
ant additional questions to answer prior to drawing a conclusion. For example, it seems to matter how the data on public services is being used. Moreover, it is not immediately clear what a discriminatory predictive risk algorithm would mean, and when this sort of discrimination is morally impermissible.

In order to consider the strength of this ethical claim against the AFST, I will first establish an understanding of what discrimination is and when cases of discrimination are wrong. Once I establish the conditions for wrongful discrimination, I will reframe Eubanks’ ethical claims of discrimination against the AFST according to this more rigorous definition, and evaluate the strength of these specific claims by considering whether the AFST fulfills these conditions. Ultimately, I will argue that the AFST does not fulfill the conditions of wrongful discrimination and therefore does not wrongfully discriminate against poor families.

1. What is Discrimination?

The term “discrimination” is used in general discourse in two main ways: (1) as a non-moralized, descriptive term or (2) as a necessarily normative assertion of disadvantage. In the first case, there is no prima facie moral judgement attached to the term, such that an act can be discriminatory, but morally permissible. In the second case, a negative moral weight is necessarily attached to the term, such that all cases of discrimination are wrong according to this definition. This second use is common in public discourse, where an assertion that an act is discriminatory implies that the speaker believes the act is necessarily morally problematic.

To spell out this distinction further, consider the following example of a prima facie morally nebulous charge of discrimination: A men’s clothing company is accused of employing only male models, explicitly using gender as a requirement for the role. One may ask two different questions about this accusation. First, whether the men’s company actually adopted a hiring policy that differentiates between workers based on sex. And second, whether this differentiation is morally wrong. The non-moralized term allows us to distinguish between these two questions. For instance, one could ask if this case is discriminatory, and if it is wrongfully discriminatory. The moralized term, on the other hand, is not nearly as precise because it captures both the act of differentiation along with the moral claim of wrongfulness.

For the purpose of clarity and precision in this essay, I will adopt the first term, with the task of qualifying whether the discrimination in question is morally problematic. Thus, discrimination will be defined in this paper as the differential treatment of an individual based on their membership in a group, defined by a specific trait. My next task is to identify the conditions under which discrimination is morally wrong.

2. When is Discrimination Wrong?

The aim in this section is to provide a general understanding of wrongful discrimination by identifying the necessary conditions that make discrimination morally wrongful. I build up a definition of wrongful discrimination, starting with the general non-moralized definition of discrimination as the differentiation of an individual based on their membership in a group as defined by a specific trait. Ultimately, I argue that discrimination is wrongful if it satisfies the following three conditions.

(1) Salient Group Membership Condition: The differential treatment of an individual based on their membership in a group defined by a socially salient trait.
(2) the Differentially Favorable Treatment Condition: The differential treatment disadvantages one group over another.
(3) the Explanatory Moral Justification Condition: There is no reasonable moral justification for the discriminatory treatment.

The following subsections further explain each condition along with an example that motivates its inclusion as a necessary condition.

2.1 The Salient Group Membership Condition

Consider the following two examples of discrimination that result in different moral conclusions. In the first example, a school teacher decides to split students into two lines for lunch based on the first letter of the student’s last name with A-M in one group and N-Z in another group. In this case, the teacher’s differentiation between students does not seem morally problematic. Now contrast this example with another school
teacher who splits his class into two lunch lines based on the student’s race, placing black and white children in different lines. The main difference between these two examples is simply the type of trait used for differentiation, which suggests that the type of trait chosen has moral weight in the wrongfulness of a discriminatory act.

The difference between the traits in the examples is that one trait refers to a socially salient group whereas the other does not. A socially salient group is one that carries important social and moral weight because the group has been historically, unjustly disadvantaged in the past. These groups generally encompass the legal category of protected classes such as race, religion, sex, age, and disabilities among others. However, the category of socially salient groups extends beyond protected classes and includes groups that have been historically disadvantaged based on economic status, prior criminal standing, or physical traits. Deborah Hellman, a legal scholar at the University of Virginia, describes the traits that define such groups as traits that have a history of being singled out for mistreatment or social disadvantage.

In the first example, the trait used to create groups for the activity is not socially salient – the first letter of a student’s last name does not carry historical moral weight, nor is it a group that is systemically disadvantaged in society. On the other hand, a student’s race is a trait that carries significant moral and societal weight, and therefore seems problematic as a use for differentiating groups. Thus, we have reason to believe that one of the necessary conditions for wrongful discrimination is the salience of the group membership. However, this trait alone is not sufficient to constitute wrongful discrimination.

2.2 The Differentially Favorable Treatment Condition

This condition seems insufficient as the sole indicator of wrongful discrimination when we look at additional examples of discrimination across socially salient groups that do not seem morally impermissible. In the first example, a doctor routinely prescribes a regular mammogram for a female patient because she is a woman. The doctor treats patients differently according to their sex. Contrast this first example from a second in which an employer decides not to employ an applicant because she is a woman. In both examples, the discriminatory agent treats an individual differently based on a socially salient group, namely sex. However, the first example does not seem like an example of wrongful discrimination, whereas the second does.

The contrasting moral intuitions in these two examples suggest that the consequences of the differentiation are important to our analysis of the moral standing of the differentiation. Specifically, it seems that the differential treatment of a discriminatory act must be differentially favorable to be considered wrongful discrimination. In the case of the doctor, there is no clear differentially favorable treatment to either group even though the differentiation happens across a socially salient group. The doctor does not disadvantage one group over another, but rather tries to offer equally favorable treatment in terms of cost and prevention to both men and women through the differential treatment given the higher occurrence of breast cancer in women. In the example of discriminatory hiring policies, there is a clear disadvantage to a historically disadvantaged group, where an applicant is denied because she is a woman. Prima facie, the hiring manager disadvantages women over men through the discriminatory hiring policy.

Based on this explanation, we can add the Differentially Favorable Treatment condition to our working definition of wrongful discrimination. This condition measures whether the act of discrimination disadvantages one group over another. With this addition, we can reason that an act of discrimination is wrongful if it disadvantages an individual based on membership in a socially salient group.

One might object to this addition by pointing to cases of wrongful discrimination where there is no explicit disadvantage in treatment. A class of examples of this type are the Jim Crow laws passed to differentiate in such a way that is separate but equal. The defense of such discriminatory laws was that there was no explicit disadvantage between the groups. This claim, however, was not true in practice, as the resources for one group were significantly worse than that of the other. However, even if it were true that resources were the same, the fact of separation itself creates a disadvantage because of the stigma and harm caused by that separation. Thus, the qualification that the discrimination disadvantages a group applies to any sort of harm or stigma as well as explicit disadvantages.
Although the Salient Group Membership condition and the Differentially Favorable Treatment condition are both necessary conditions in a definition of wrongful discrimination, they are together insufficient to constitute wrongful discrimination.

2.3 The No Independent Moral Justification Condition

The above definition of wrongful discrimination seems insufficient when we consider examples of discrimination where a clear disadvantage exists across a socially salient group but the action is considered morally permissible. For example, suppose the state denies a twelve-year-old a driver's license. In this example, the state disadvantages a person through withholding the privilege to drive because of the person's age, a protected class. According to our current definition, this should be a case of morally impermissible discrimination, yet moral intuition seems to say that this case is permissible. This case is permissible because the differentiating trait used by the state — age — directly explains the differential treatment in this case. That is, the socially salient trait of a driver's age offers a direct, morally plausible justification for the disadvantage through the presence of a clear explanatory link between the trait and the differentiating act.

The purpose of this first example is to establish the basis for a plausible, explanatory moral justification. It is possible to offer a reasonable justification for a clear disadvantage to a socially salient group that would render the discrimination morally permissible. The standard of reasonableness for a given justification requires a context-specific analysis of the explanatory strength of the justification. Thus, we can add the Explanatory Moral Justification condition to our working definition but will also consider more examples to understand the limits of reasonableness for these justifications.

In order to understand the limits of a justifiable explanation, I will consider additional examples of discrimination where multiple justifications are offered. For example, consider a construction site that decides not to hire an applicant based on the existence of a criminal history of a non-violent misdemeanor, a shoplifting charge, that took place over a decade ago. Contrast this example from one in which a daycare decides not to hire an applicant based on a criminal history that involves a similar category of gross misdemeanor as the first, except the specific charge is one of child abuse. In both cases, the category of an ex-offender constitutes a socially salient class and the ex-offender in this case is disadvantaged over other job applicants. However, the first seems like wrongful discrimination whereas the second seems morally permissible.

Both examples are cases of discrimination where a person is disadvantaged due to membership in a socially salient group. However, the justifications given are not equally reasonable. In the first case, the ex-offender's criminal record is not directly related to the job in question. Namely, a prior shoplifting charge is not directly related to an applicant's ability to work in construction. There is no direct explanatory link between the disadvantage and the reason given. The lack of a proper justification suggests that this first case is one of wrongful discrimination.

On the other hand, the justification given by the employer at the daycare is reasonable, in that the ability to work well in a daycare requires that the applicant can work well around and with children. The existence of a substantiated charge of child abuse is directly related to an applicant's ability to work at a daycare because it suggests that the applicant is unable to properly carry out this job. The justification given is a reasonable explanation for the disadvantage which indicates that this act of discrimination is not wrongful.

Thus, the final account of wrongful discrimination is a form of differentiation that disadvantages based on a socially salient trait without a justifiable explanation for the disadvantage. An act of discrimination is wrongful when it satisfies the following three conditions: (1) the Salient Group Membership Condition, (2) the Differentially Favorable Treatment Condition, and (3) the Explanatory Moral Justification Condition. Based on these conditions for wrongful discrimination, we can consider whether the AFST wrongfully discriminates. The next section focuses on objections to the claim that the AFST can be discriminatory, which will motivate a distinction between two types of discrimination, direct and indirect.

3. Types of Discrimination: Direct vs. Indirect

Despite the establishment of conditions under which
discrimination is wrongful, it is not immediately clear what it would mean for a predictive risk tool like the AFST to discriminate. One might object to the general claim of discrimination against the AFST, arguing that only people can discriminate, not algorithms or software systems. Each of the examples of discrimination offered earlier in the paper concern people discriminating against other people without machines. Moreover, one can claim that predictive risk assessment systems are objective in a way that humans generally are not.

There are two ways in which a predictive risk assessment system like the AFST can be discriminatory. First, the humans who created the AFST could use the tool as a means to discriminate against others, either explicitly or while hiding behind the veneer of technological objectivity. The AFST, as with all other risk assessment systems, was created by humans who potentially have biases or intentions to disadvantage a group. The creators and users of the AFST could use the tool to discriminate against certain people.

Second, the predictive risk assessment tool itself could make societal decisions that result in the differentially favorable treatment of a socially salient group, even if there is no intent by the creators to discriminate. Understanding the AFST as a decision-making agent that acts through its release of predictive scores for individuals and families, we can see that the decisions made by the AFST could systematically disadvantage one group. Even though these software systems can be more objective than humans in their analysis, they often still capture the underlying biases that exist in the environment, even if the creators did not intend for the tools to do so. This happens in many of the creative steps involved in the construction of these systems, such as data collection, data training, and the choice of predictive variables among others.31

Based on these observations, one way to determine whether or not the AFST is actually discriminatory is to focus on the intentions and biases of its creators and users. If the creators and users have a clear bias or intent to discriminate against a certain socially salient group, then there is reason to believe that the AFST is discriminatory. If there is no clear bias or intent to disadvantage, the AFST can still be found to be discriminatory through the existence of an unintended disadvantage to a socially salient group. These two types of discrimination fall along the philosophical distinction of direct and indirect discrimination.32 In this section, I will distinguish between these two types of discrimination in order to ultimately argue that Eu-banks’ claim of discrimination is likely one of indirect discrimination.

In the direct case, the discriminating agent intentionally treats a group of people worse than another because of a bias against that group. This is the more paradigmatic example of discrimination, in which the differentially favorable treatment is motivated by animus. Consider the example of a restaurant owner who denies service to black guests because the owner does not like black people. The owner intentionally and explicitly treats black guests differently and worse than white guests because of a specific trait, their race, resulting in disparate treatment of black guests. I argue that the AFST is not directly discriminatory. Both the DHS and Vaithianathan’s team are clear in their concern about discrimination and their desire to mitigate it, suggesting that they have no intent to discriminate nor any desire to promote differentially favorable treatment.33 Moreover, the creators of the system approached the development of this tool with a high level of transparency, indicating that they are not concealing any discriminatory intentions.34 Based on these observations, it seems that the AFST is not discriminatory with regards to an existing intent to discriminate.

Absent any intent to harm or bias against a group, an agent can still wrongfully discriminate against a group if the agent’s actions result in a disproportionately negative effect on that group without reasonable justification. Consider the paradigmatic example of the U.S. Supreme Court decision in Griggs v. Duke Power (1971), in which a utilities company in North Carolina decided to use written tests to determine promotions, resulting in an almost full exclusion of African American employees from promotion. The company was not accused of directly discriminating against African Americans, but given the disproportionately disadvantage to this socially salient group coupled with a lack of clear reason to use the written test, the U.S. Supreme Court considered this wrongful indirect discrimination.35 Without a reasonable justification for the existence of this disadvantage, the unfavorable differentiation is considered wrongful. Unlike direct discrimination, indirect discrimination does not require an intent or motivation. This type of
discrimination depends on the existence of differentially favorable treatment. Tools like the AFST can be indirectly discriminatory in a number of ways. For example, the risk scores given by a tool could vary on average across a socially salient group. Or, it could turn out that the risk scores given to one group are systematically less accurate than those given to another group. Both of these seem like ways in which a tool like the AFST can be indirectly discriminatory.

In the following parts of this chapter, I will consider whether Eubanks’ claim of indirect discrimination fulfills the conditions of wrongful discrimination developed earlier. This will entail developing a more precise version of Eubanks’ argument based on the direct and indirect distinction followed by an analysis of the wrongfulness of her charge of indirect discrimination. This analysis will proceed through a progression of the three conditions of wrongful discrimination considered earlier: (1) the Salient Group Membership condition, (2) the Differentially Favorable Treatment condition, and (3) the Explanatory Moral Justification condition.

4. Eubanks’ Charge as Indirect Discrimination

Based on the earlier discussion, we can formulate Eubanks’ charge as one of indirect discrimination in which she argues that the AFST systematically fails the poor. She claims that “the AFST interprets the use of public resources as a sign of weakness, deficiency, even villainy.” Her claim rests on a belief that a family’s AFST score is higher with the inclusion of the use of public services as a feature than it would be without its inclusion as a feature. This initially seems problematic given that public services are predominantly used by those with a low socioeconomic standing which may indicate differentially favorable treatment. Eubanks claims that the AFST scores calculated for poor families are generally higher than those for a middle-class or wealthy family because of their socioeconomic status.

Eubanks establishes her claim by pointing to the high number of predictive variables in the AFST that are direct or indirect measures of poverty. Of the 131 predictive variables in the AFST published by the DHS in its efforts at transparency, almost half are either direct or indirect measures of poverty. For example, a quarter of the variables track direct measures of poverty like the “use of means-tested programs such as TANF [Temporary Assistance for Needy Families], Supplemental Security Income, SNAP [Supplemental Nutrition Assistance Program], and county medical assistance.” Another quarter of variables track indirect measures of poverty like “interaction with juvenile probation.” Based on this evidence, Eubanks constructs a charge of wrongful discrimination against the AFST.

Prima facie, it seems that Eubanks’ charge fulfills the conditions of indirect wrongful discrimination. The Salient Group Membership condition is fulfilled as the differential treatment occurs for poor families, a socially salient group with a history of disadvantage. Moreover, an initial pass suggests that the differential treatment disadvantages the poor. According to Eubanks, we have reason to believe that the risk scores for poor families are higher on average, which means that poor families are investigated more often. Eubanks draws on anecdotal evidence from families that have experienced fear, intrusion, and loss of privacy sometimes caused by these investigative visits from government officials. Based on Eubanks’ findings, it seems that increased investigations can constitute real harm. Therefore, it seems that the Differentially Favorable Treatment Condition can be satisfied.

Finally, we must consider the Explanatory Moral Justification Condition. In order to satisfy this condition, there must be an absence of a reasonable moral justification for the existence of the disadvantage outlined above. At first glance, it seems that one reasonable justification for the existence of a higher risk score for poorer families is that poorer families are indeed at higher risk of child abuse than their economically more fortunate counterparts. In other words, it may be the case that the base rates of child abuse for poorer families are actually higher than those of richer families meaning that the risk assigned is accurate. Given it has empirical backing, it seems to be a reasonable justification. The existence of a higher risk of child mistreatment for poorer families seems to initially justify the issuance of a higher AFST risk score.

Based on this initial pass through the three conditions, it seems that Eubanks has a plausible charge of wrongful indirect discrimination against the AFST. In the following section, I consider a defense of the AFST from the DHS by outlining the progression of the dialogue between the DHS and Eubanks, looking
at both the DHS’s initial response to Eubanks followed by Eubanks’ follow up response to the DHS. Then I will evaluate the strength of these responses, and ultimately argue that the DHS’s response to Eubanks’ charge of indirect discrimination is insufficient because it does not directly address Eubanks’ claim.

5. The Public Debate: Eubanks vs. DHS

In response to Eubanks’ charge of discrimination, the DHS produced an executive statement in which the DHS argues that the AFST does not actually disadvantage poor families, and is therefore not wrongfully discriminatory. The DHS points out that a family’s use of public services in Allegheny County has a positive correlation with the likelihood of child abuse. That is, the use of public services actually decreases the scores for families that access the public services. In their executive statement, the DHS cites the statistic that “in reality, for 45% of families, receipt of the services is protective, that is, their receipt lowers the AFST score.” This means that the use of public services actually has a negative correlation with the AFST score for a little under half of the families. A family’s AFST score is often lower if they use public services than if they do not.

Eubanks responds to the DHS by disputing the claim that the use of public services is protective. She focuses on the same statistic that the DHS relied on and suggests that a score that is protective for only 45% of families is not actually protective for the majority. On her website, Eubanks responds to the DHS’s executive statement by arguing that “the County’s January 31 statement suggests that for 55% of families – the majority – receipt of public services does in fact raise their AFST score, leaving them disproportionately vulnerable to child welfare investigation.” Eubanks attempts to use the complement of the same statistic given by the DHS to argue that the use of public services in the AFST score leads to higher scores for the majority of families.

Eubanks’ response is unsuccessful because her analysis of the data overlooks other important possibilities. We cannot assume that the use of public services raises the AFST score for 55% of families based on the fact that it is protective for 45% of families. From this statistic, we only know that the feature is protective for 45% of families, which means that the feature has either no effect or a negative effect on the remainder. Of the remaining 55% of families, it could be the case that the use of public services as a feature has a negligible effect on the AFST scores. Even though Eubanks’ follow-up response is unsuccessful, her charge of wrongful discrimination against the AFST is not affected by the DHS’s response.

The response by the DHS is insufficient in addressing Eubanks’ claim of discrimination because the response subtly answers a related, but distinct question from the one that Eubanks’ claim poses. Eubanks argues that the inclusion of the use of public services as a feature in the AFST leads to higher scores than if the features were left out. However, she is not arguing that the actual use of public services has no benefit for families who need it. In fact, Eubanks would likely agree that a family’s use of public services is protective for the family because the family receives important and helpful services such as childcare assistance or essential mental health services. While a family’s use of public services may be protective, the inclusion of that fact in the AFST may not be. The AFST’s inclusion of the fact that a family uses public services might actually increase a family’s score when compared to families who have accessed similar private services but for whom this fact is not recorded.

The DHS focuses on the claim that a family’s actual use of public services leads to lower scores on the AFST compared to the same family had they not used public services. Eubanks, on the other hand, focuses on whether or not the AFST’s inclusion of the fact of the use of public services increases AFST scores compared to the same family had they used private services instead. The DHS does not address the primary question of how the actual inclusion of the fact of the use of public services affects a family’s score. This is an empirical question that would require additional research in order to reach a conclusion, but Eubanks’ claim seems plausible.

For now, I will assume that the inclusion of the predictive variable does in fact increase scores for poor families. That is, I accept the claim of differential treatment across a socially salient group for the sake of argument. Even with this assumption, I argue that a higher AFST score does not necessarily mean a disadvantage to the family with the higher score, and therefore, there is no clear differentially favorable treatment across a socially salient group. I will construct
this argument by first analyzing the moral effects of a high AFST score. This moral analysis will consider the effects of investigations on the different moral stakeholders.

6. Moral Analysis of the Effects of Investigations

Contrary to Eubanks’ assumption, a higher AFST risk score and a subsequently higher rate of investigations, on its own, may not necessarily be more harmful than a lower score and a lower rate of investigations. Eubanks’ reasoning is as follows: higher AFST scores lead to a higher number of investigations. Investigations are generally harmful to families because of their intrusive nature. Thus, higher AFST scores are generally more harmful to families than lower AFST scores. Based on this reasoning, the moral weight of the risk scores depends on the harmfulness of investigations. If it turns out that investigations are not more harmful, then higher AFST scores for one group would not constitute differentially favorable treatment. Thus, the moral analysis in this section will focus on the effects of government investigations.

The moral weight of these investigations depends on the outcome of the investigation and the moral stakeholders affected by the decisions. Eubanks correctly points out that these investigations can be harmful to families through an intrusive investigation, a loss of privacy, and potential repercussions for parents. However, investigations can also lead to important resources for parents, increased stability in the community, and most importantly, safety to a child who was in an abusive situation. Thus, the moral effects of an investigation seem to depend on the ultimate outcome of the investigation, whether or not the investigation leads to significant findings.

An analysis of the moral stakes of accuracy depends on an understanding of the benefits of a correct investigation and the harms of an incorrect investigation. The moral analysis in this section will focus on the benefits of true positives and true negatives, as well as the harms of false positives and false negatives for different stakeholders with partially aligned interests. This will entail an initial discussion of the composition of the main moral stakeholders in an investigative case, followed by the effects of a true positive, true negative, false positive, and false negative on these different stakeholders. After this moral analysis, I will consider the question of whether Eubanks’ charge of a higher risk score for poor families disadvantages them. Eubanks’ claim focuses on the disadvantage of scores to families as a singular unit. However, families are constructed of individuals that could have different moral claims and needs. For example, the parents might have a claim to custody of their child that conflicts with a child’s claim to a non-abusive living environment. With regards to an investigation, the three main moral stakeholder groups within a family are the parental guardians, the children suspected to be at risk of abuse, and other children or extended family. Each of these groups has potentially competing claims on an investigation of child abuse directed at the family. The children have a moral claim to safe living conditions that are free of abuse or extreme neglect. The parents have claims to a degree of privacy and autonomy in their own child rearing practices as well as a claim to custody of their child. The remaining family members have claims to privacy in the generally private spheres of life within a home. Each of these claims works together to build some of the moral landscape that a decision to investigate a family for potential child maltreatment must consider.

6.1 True Positive and Negative

First, a true positive in the AFST means that a high risk score is correctly given, meaning that an investigation was substantiated through action. These actions fall into two general categories. In the first case, an investigation can lead to the placement of a child into foster care. This happens when the Child Protective Service (CPS) caseworker who conducts the investigation assesses that the living situation constitutes child abuse according to the state’s legal definition of child abuse. Another actionable outcome of an investigation is the provision of identified necessary resources for the family to reduce risk of harm to the child. The investigation also involves help in identifying goals and action steps for family members as well as a safety plan.
leads to action, the benefits of the investigation can be significant.

On the other end, a true negative in the AFST means that a low risk score is correctly assigned to a family, meaning that families who do not require investigation are left alone. Through powerful anecdotes, Eubanks reveals that investigations can be harmful to families, and avoiding these investigations when they are not necessary is an important benefit. Investigations can take several weeks and include unannounced home visits, can include interviews with different family members and members of the community, and can carry negative stigma within a community. As such, investigations can be intrusive and the avoidance of unnecessary investigations is an important benefit. Both the true negatives and true positives provide important benefits to the individuals involved, such that having fewer of these benefits can constitute a harm. If the AFST is less predictive for one group than another, then that group misses out on important benefits that come with a properly run child welfare system. The benefits of true positives and negatives are tied to the harms associated with an investigation for individual families. Therefore, an investigation that turns out unsubstantiated harms a family as much as a true negative benefits them, and a lack of an investigation when conditions of abuse exist harms the family as much as a true positive benefits a family.

6.2 False Positive and Negative

A false positive indicates an incorrectly high AFST score, which means that it is more likely that an investigation takes place in a family where no conditions of child abuse exist. A false investigation is generally harmful to all of the relevant stakeholders within an individual family. For parents, an investigation can carry with it an invasion of privacy, an increased negative stigma from the community, the potential loss of custody over a child, fear of future retribution, and feelings of being targeted tied to a loss of security. For the children, the potential harms of an investigation include the fear of future loss of parents, increased stigma from peers or neighbors, and the potential harms involved with being taken away from parents. One might object that the collective pursuit of investigations can justify these mistakes because of the importance of keeping children in the community safe. However, the moral analysis here is only focused on the consequences of a single investigation on the direct moral stakeholders affected, rather than the overall moral permissibility of having investigations in general. The aim here is not to argue that investigations are harmful in general, but rather that a wrong investigation is harmful to the family that received the false positive score.

As with false positives, false negatives also lead to harmful consequences for the stakeholders involved. A false negative indicates a mistakenly low score for a family, meaning that families that have unsafe, potentially abusive living conditions for the children are passed over for investigation. The clearest harm in this case is to the child who is left to live in an abusive living situation without mediation. This is a clear harm that takes precedence over some of the harms of having the investigation itself. Moreover, there is an additional harm to parents who may not receive the support they need from CYF if the family is mistakenly passed over for investigation.

7. Evaluation of Disadvantage

Based on the moral analysis above, I will argue that there is not a clear disadvantage to poor families based on the differential treatment of scores. That is, higher average AFST scores for poor families do not constitute differentially favorable treatment. A higher average AFST score means that a family is more likely to be investigated. A higher likelihood of investigation, however, does not constitute a clear disadvantage. The preceding moral analysis uncovered the nuanced moral weight of investigations, where true positives and true negatives generally benefit a family and false positives and false negatives generally harm a family. If it is the case that the greater number of investigations of poor families are substantiated at equal or greater rates, then these additional investigations might actually benefit rather than harm the families. Whether or not it is a benefit or a harm depends largely on the accuracy of those investigations.

7.1 Disadvantage Through Differences in Accuracy

Thus, it seems that Eubanks’ initial charge of indirect discrimination does not necessarily disadvantage the poor, and is therefore not necessarily discrim-
whether the investigation disadvantages the family depends on the accuracy of the investigation. To further examine whether the AFST wrongfully discriminates against poor families, we can consider instead a modified charge of discrimination focusing on a difference in levels of accuracy rather than scores. I will first consider whether there could be differential treatment across accuracy in the AFST. For this, we will look at another step in the development of a predictive model that is susceptible to capturing human bias, namely the training data. The training data is the initial data from which correlations are drawn to the desired outcome variable. This data should be representative of the population for which the predictions from this data is being made.

If the data collection is skewed towards a specific subset of the population, then the accuracy of predictions for this subset may be different than for those of the rest of the population. In the case of the AFST, Eubanks’ presents a real concern that the data is misrepresentative, gathering additional data on families from a lower socioeconomic status. This concern stems from the fact that some of the data points gathered include direct and indirect measures of poverty, as discussed earlier. It does seem that there is simply more data on poor and working-class families than upper-class families given that many of the data points measured are from the use of public services.

As in the previous section, an empirical question remains to show whether there is a true difference in the accuracy of the tool across socioeconomic status. However, Eubanks’ account gives us reason to believe that this might be the case. However, a greater amount of data on families from lower socioeconomic backgrounds might actually lead to greater accuracy for this group than for other groups. Given that there is more predictive data on poor families to make statistical correlations from, the accuracy of the predictive risk scores for poor families might be greater than for those with less meaningful data.

If this is the case, it seems that there may not be a disadvantage to poor families according to accuracy. Based on our earlier moral analysis, we determined that the accuracy of the machine has moral salience, with a greater accuracy constituting more benefits, and a lower accuracy indicated more harm. Thus, there does not seem to be a disadvantage to poor families, but rather a potential advantage through a higher proportion of true positives and true negatives for this socially salient group.

7.2 Disadvantage Through Differences in Error Rates

However, one might object that even if the accuracy of risk scores is calibrated across socioeconomic status, a difference in rates of error between the two groups may lead to a meaningful disadvantage. To see how this might be the case, consider the following example of a Propublica investigation of a risk assessment tool for determining the likelihood that a criminal might recidivate, called COMPAS. The tool is intended for use by judges in court cases to assist in sentencing decisions.

This analysis of ProPublica found that, while the tool was equally accurate in its assessment of risk for both black and white criminals, the distribution of error types were skewed across race in a way that constituted a clear disadvantage for black criminals. Even though the tool was equally accurate for both groups, meaning that the same proportion of mistakes were made for both groups, the types of errors made for each group were importantly different. There are two types of errors that can be made: false positives and false negatives.

In the case of the COMPAS algorithm, a false positive is a score that mistakenly flags an individual as being at high risk of recidivism when in reality the individual is at a low risk. On the other hand, a false negative is a score that mistakenly flags an individual as being at low risk of recidivism when in reality the individual is at a high risk. Both mistakes constitute an error, however each type of error holds importantly different consequences for the individual. A false positive can be very harmful for an individual, potentially leading to a higher bail posting than is necessary. However, a false negative may actually be beneficial to an individual, leading to lower bail postings than is necessary for the criminal.

The analysis found that the distribution of errors was such that criminals who are black experience a significantly higher proportion of false positives than false negatives whereas white criminals experience more false negatives with fewer false positives. This seems like a clear disadvantage to a socially salient group even though the tool is equally accurate for both groups. Similarly, one might object that the
AFST might disadvantage a socially salient group not through different levels of accuracy, but rather through different distributions of error rates.

Unlike the COMPAS example, however, both false positives and false negatives are harmful to the relevant stakeholder with regards to the AFST. This is the case because the relevant moral stakeholder in the case of a CYF investigation, a family under investigation, contains multiple potentially competing interests, such that the harms and benefits of a false positive and false negative are more complicated. In the case of the COMPAS algorithm, an individual is the only relevant stakeholder, so one type of mistake is more clearly harmful than the other.

Based on the moral analysis in section six, both false positives and false negatives are harmful to the relevant stakeholder. It is also unclear whether one type of harm is necessarily worse for a family than the other, as can be seen through the ongoing debate in philosophical literature. As a result, even if the AFST does have skewed error rates across socioeconomic class in one direction or another, it does not seem that this constitutes a clear disadvantage for one group over another.

Conclusion

I have argued that the AFST is not wrongfully discriminatory. The AFST is likely not directly discriminatory given the lack of evidence of any ill intent from the creators of the machine. Moreover, the two charges of indirect discrimination seem not to fulfill the second condition of wrongful discrimination, namely, the Differentially Favorable Treatment Condition. In conclusion, it seems that Eubanks’ claims of wrongful discrimination against the AFST are unsubstantial, and that the tool is not wrongfully discriminating against poor families.

Endnotes

2. Vaithianathan et al., “Developing Predictive Models to Support Child Maltreatment Hotline Screening Decisions: Allegheny County Methodology and Im-

plemnetation.” Center for Social data Analytics (April 2017)
4. Ibid, 46.
6. In limited cases an additional field screen may be conducted in the call screening phase prior to making the decision to screen-in or screen-out for a formal investigation.
11. The AFST previously consisted of two underlying models. But, with the publication of the methodology for Version 2 of the AFST in April 2019, one of the models, the re-referral model, was removed leaving only the placement model.
14. Ibid., 11.
15. Ibid., 15.
17. Ibid., 2.
18. Ibid., 2.
20. This argument was offered prior to the release of the impact study that found marginal increases in one measure of accuracy and marginal decreases in another. This paper was published in April of 2017 whereas
the impact study was published in April 2019.


23. Ibid., 118-122.


26. This example is an adaptation of an example from the first chapter of Hellman, Deborah, *When Is Discrimination Wrong?*

27. Philosophers like Deborah Hellman, Owen Fiss, and John Hart Ely agree that the type of trait chosen is important to an understanding of the wrongfulness of discrimination. Specifically, they agree that traits that point to historically disadvantaged groups carry moral weight in making discrimination wrongful. However, much of the dialogue between these philosophers is around the question of why this condition is important to discrimination, with the assumption that it is important. For the purpose of this paper, we can leave the question of why this condition is important aside, as we only need to know what the conditions are for wrongful discrimination.


29. There is a similar example in Hellman, *When is Discrimination Wrong?*

30. The least serious cases of child neglect are considered gross misdemeanors in the same category as petty theft or driving under the influence of alcohol.


37. Ibid., 125


39. Executive Statement from County of Allegheny, County Executive Rich Fitzgerald

40. There are other moral stakeholders affected by the decision to investigate a family, including the surrounding community, the investigators themselves, and the child protection services. But, this analysis focuses only on the family unit, as this is the basis of the claim of harm proposed by Eubanks.

41. According to the Pennsylvania Family Support Alliance, child abuse is “when an individual acts or fails to prevent something that causes serious harm to a child under the age of 18. This harm can take many forms, such as serious physical injury, serious mental injury, or sexual abuse or exploitation.” PA Family Support Alliance. “Abuse and Neglect Definition” https://www.pa-fsa.org/Mandated-Reporters/Recognizing-Child-Abuse-Neglect/Abuse-Neglect-Definition


43. Some might dispute this indication of harm by arguing against the child welfare system and the government's role in ensuring the safety of children in general. This is another lively debate, and one that I will set aside.

44. The investigation might be beneficial to other actors, like the state or the community through the knowledge that there is no child abuse. But, for the purpose of this section, we are only looking at the benefits and harms to the specific family affected.

45. There are vibrant debates in related literature about the nature and extent of the harm of taking a child away from a parent, and whether this is justified at all, let alone which conditions it is justified in. These discussions focus on the ethical challenges of investigators, but for this paper, we can lay this debate to the side.


Bibliography


Stevenson, Megan, and Jennifer L. Doleac. “Algorithm-


Introduction

Discourses around fitness and health, specifically defining “health,” how to pursue it, and how to measure success have long been prevalent and highly debated topics in society. Health and fitness ideals have become increasingly hegemonic in U.S. society as concerns rise in relation to the “obesity epidemic.” Although institutions reframe these ideals and messages over time, ultimately the fitness, diet, and media industries will always emphasize the idea that losing weight and obtaining an idealized body type is good for one’s health. This debate assumes that weight itself is a stand in for health, and that if an individual is overweight they must be unhealthy, and if an individual is “normal” they must be healthy.¹

Social systems including the media have shifted the discussion over time towards health and wellness rather than thinness; however, a thin and toned body type still remains the underlying ideal for most. In a study of the “fitspiration” movement on social media, meaning the promotion of healthy eating, exercise and self-care rather than thinness and weight loss, the scholars’ content analysis still found that one body shape was overwhelmingly represented in the media.² In line with their hypothesis, this was a fit but also thin and toned body, in accord with current sociocultural trends.³ Reframing the discussion and shifting...
the type of idealized body does not free individuals from self-scrutinization but continues to oppress both men and women to obtain the ideal and purchase goods and services from the very institutions shaping this discussion. Capitalist profit, societal norms and “fat phobia” are not simply pushing individuals to be “healthy,” but at a deeper level these mechanisms convince us that to be truly healthy and happy citizens we should diet, exercise and try to lose weight to achieve the idealized body.

College students specifically seem to be concerned with weight, body image, and fitness. Female college students in particular report a high level of body image and body satisfaction problems. Certain body attitudes can be tied to peer-dominated atmospheres, like college campuses, where students are exposed to a high level of social comparison. This study seeks to understand the ways in which college students regulate the self in order to achieve a beauty and body standard perpetuated by society and heightened through an atmosphere of continual peer comparison. Our study seeks to understand the experiences of students who utilize market solutions in order to regulate and maintain their body to achieve the norm.

In recent years, tracking devices and wearable sensors have emerged as a key mechanism in the construction of the healthy and responsible citizen. The rise of digital technologies in the fitness sphere allows individuals to produce quantified measurements of one’s body, such as heart rate, caloric expenditure, or steps taken per day. In turn, the data generated allows users to collect information about themselves, and reflect on how this information can be used to improve their lives in some way. The use of fitness hardware and software as a form of production and consumption of one’s own data is known as “self-tracking” or “reflexive self-monitoring.”

Studies have shown that a high percentage of individuals across all age groups use tracking as a way to measure health. 60 percent of US adults are currently tracking their weight, diet, or exercise routine. Older adults are more likely than younger ones to track their weight, diet, or exercise routine: 71 percent of those ages sixty-five and older do so, compared with 61 percent of 18 to 29 year olds. In addition, the Pew Research Center finds that 68 percent of college graduates track their weight, diet, or exercise routine. Therefore, this article addresses how college students grapple with and make meaning in the context of fitness tracking technology, as the phenomenon is increasingly prevalent among people in this age group.

Specifically, we explore how students from a selective Northeastern liberal arts college experience the “datafied self” in terms of health behaviors, agency, and emotions. The “datafied self” refers to the ability of an individual to see themselves, such as their activity and their body movements, through a digital interface. Many studies have addressed the ways health and fitness apps operate, how digital technologies influence our understanding of health and fitness, and what information fitness technologies aim to elicit from users. What is missing, however, is a longitudinal examination of health approaches before and after using fitness tracking technology. Therefore, this research builds off of previous studies by comparing experiences and health behaviors before and after using a fitness tracking device. For example, we explore whether college students feel more responsible to pursue their health goals when using the device. Furthermore, we examine agency surrounding fitness tracking technologies, including critical perspectives, decisions to turn them off, or ignoring the device and its reminders.

Wearable devices and apps are purposefully used as activity monitors or fitness trackers. These devices and apps track everyday activities, and help individuals achieve health-related goals such as weight loss or regular exercise. One study claims that the design of digital tracking technologies “builds on the notion that visibility is desirable and that it is of value for people to have their physiologies and everyday movements made observable and legible.” Fitbit’s company website states: “Fitbit’s accelerometer constantly measures the acceleration of your body and algorithms convert this raw data into useful information about your daily life, such as calories burned, steps, distance, sleep quality.” According to a study by the Pew Research Center, 46 percent of trackers say that self-tracking has changed their overall approach to maintaining their health or the health of someone for whom they provided care.

Our research takes place at a private liberal arts college, where students are highly competitive, predominantly white, and mostly come from upper to middle class backgrounds. One study regarding the medicalization of college students at a similar institution of higher education finds “these students are in advantageous positions; their economic resources have
given them the ability to seek out physicians, obtain medical treatment, and engage in consumption based body projects.”

The context of students at this college campus is important for our research as the institution provides access to a number of athletic facilities, personal trainers, free access to the gym, free workout classes, wellness institutes, and dietitians for students. Access to these resources and access to fitness trackers excludes individuals who do not have access from our study. In addition, previous student research at this institution has found that “being fit is an important part of the students’ habitus… they see going to the gym as a valuable behavior.”

Therefore, the intense fitness culture as well as access to many health and fitness resources at this institution of higher education, much like similar competitive institutions, provides a research environment where optimization of the body is highly prevalent. As a result, we examine how fitness tracking plays a role in optimization.

Literature Review

Foucault’s “technologies of the self” and biopower

Many scholars have linked neoliberal values to the rise of individual pursuits of self-optimization in relation to health. Rather than relying on institutions of power to institute coercive measures, neoliberal political systems invest faith in the voluntary take-up of mandates by citizens themselves. Health has become one area in which citizens are expected to voluntarily “take-up,” placing responsibility on the individual. Crawford coined the term “healthism” to describe a discourse in public health practice in which individuals are held responsible for the prevention of illness by knowing and avoiding the risk factors associated with ill health.

As a result, care of the self has become an ethical project: individuals are expected to acquire knowledge, think critically and reflect on health-related information, and thus adopt self-awareness of how to regulate and maintain health. An ethical project implies that one chooses whether or not to pursue a certain lifestyle: it is up to the individual to pursue a life that is healthy or not, a body that is fat or not. People who take on this sense of responsibility for their health are moral and good, and those who do not are lazy and irresponsible. Additionally, framing health through an ethical lens allows a neoliberal society to dictate whether certain health approaches are considered “good” and which are considered “bad,” and also which market solutions will help put us on a straightforward path to health.

Individuals engage in the pursuit of health through purchasing goods and services that will allow them to achieve their goals. The fitness and dieting industries make billions of dollars every year by creating a new product, app, fitness tracker or other quick way to lose weight and get healthier. According to a market research report, the global weight management market was valued at approximately USD 212.1 billion in 2018 and is expected to generate around USD 348.1 billion by 2025. Since weight is assumed to be an indication of one’s health, corporations profit from the idea that individuals need to do whatever they can to achieve a normalized and idealized body type. While maintaining a healthy weight is linked to lowering the risk of many types of diseases and cancers, these industries often push us to buy quick solutions or diet plans that don’t work in the long run, creating a consumer cycle where dieters and health pursuers come back again and again for another quick fix. Since society frames health as the responsibility of the individual, people are expected to navigate the confusing waters of the health and fitness industry, purchasing goods and services along the way to try to figure it out. Those that do not have economic means or access to education or resources to embark on a health journey often are not able to pursue the “healthy” and “good” citizen ideal, but through no fault of their own. Additionally, even for those who strive for these ideals, such a profitable industry pushes us to never be fully content with ourselves, and to purchase goods and services as a way to self-optimize, which ultimately generates more revenue.

Our research looks at fitness tracking devices, one of the many market solutions created by the health and fitness industry to encourage people to lose weight and achieve a normalized body type in an era of neoliberalism. Therefore, this research uses a neoliberal framework in order to understand how market solutions reinforce the idea that health is the responsibility of the individual, specifically looking at health behaviors and emotions in this context. In addition, this framework helps us to understand how users of self-tracking devices may feel heightened pressure to work out and engage in projects of self-optimization.
because they are using a device that is a physical and constant reminder to pursue healthism.

The rise of digital technologies and the focus on personal responsibility in a neoliberal context disregards the social and structural barriers to achieving health norms and standards. Market commodities such as self-tracking devices and health and fitness apps are posed as solutions to public problems such as obesity, illness, and poor health. Rather than acknowledging the social and structural factors that greatly influence health, such as low socioeconomic status, race, geographic location, or gender, discourses around health today perpetuate the idea that health is equally attainable across all demographics. Furthermore, these technologies perpetuate a “self-help” discourse, meaning that goals do not come naturally, but must be consistently worked towards with the assistance of an expert or through purchasing items to achieve one’s goals. The inability to achieve these aspirations and overcome difficulties becomes recognized as the fault of the individual rather than of their relative social and economic advantage.

Foucault provides a framework for understanding how individuals use fitness technologies as a form of self-optimization. Foucault’s “technologies of the self” are the ways in which people “engage in practices which permit individuals to effect by their own means, or with the help of others, a certain number of operations on their own bodies or souls, thoughts, conduct, and way of being, so as to transform themselves in order to attain a certain state of happiness, purity, wisdom, perfection, or immortality.” Foucault argues that power dynamics and social structures play a role in a citizens’ pursuit of self-betterment: “the manner in which one ought to transform oneself as an ethical subject acts in reference to the prescriptive elements that make up the code.” Foucault’s “code” can be applied to discourses in healthcare, politics, and the media that perpetuate an ideal slim, toned, and fit body among society today, placing responsibility on the individual to achieve this body type.

Foucault’s concept of biopower identifies how institutions maintain power through promoting normative expectations surrounding the ways in which citizens should conduct their lives. Biopower developed as a way to discipline the individual body through the governing regimes of various political institutions including the prison, school, and factory. Through “continuous regulatory and corrective mechanisms” within society, institutions of power monitor individuals into docile bodies, ultimately normalizing continuous pursuits of self-betterment. The term “docile bodies” refers to the way in which institutions discipline citizens and set expectations to maintain order—people are willing to change themselves to meet norms and standards. In the context of health and fitness, docile bodies reflects people’s inclination to change the visual appearance of their body to meet expectations set by societal institutions.

Fitness tracking devices reflect the use of biopower by our institutions. These devices have become “corrective measures” used by individuals to monitor their own bodies in pursuit of self-optimization and achieve an ideal body type. These devices are corrective measures because they provide information about ourselves, and inform us how we can improve upon that information and data to lead healthier and better lives. In addition, corporate institutions design and sell these products to normalize and correct deviance, such as an overweight or obese body.

While Foucault’s “technologies of the self” provides a framework to understand how control over bodies takes place at a macro-level, stemming from Foucault’s work on biopower comes the concept of biopedagogy, which relates to how citizens learn the norms and expectations set by institutions of power. This concept helps us understand how control over bodies also takes place at a micro-level, or through teaching and training the individual. The concept of biopedagogy describes the processes of learning and training bodies how to live, particularly how institutions of power regulate life and teach us how to live, how to move or how to look. Scholars argue that biopedagogy in the context of wearable tracking devices results in forms of training and knowledge production through “micro-practices,” which create new meanings about the technology as well as new identities. Examples of micro-practices are the device’s reminders to walk a certain number of steps in an hour or make small changes in one’s lifestyle through using a self-tracker. Scholars argue that individualized forms of biopower emerge through rising concerns over the “obesity epidemic,” which promotes moral imperatives to be healthy. Society has normalized the process of tracking and monitoring the self as a form of optimization. As such, our research explores how bodies are
technologically surveilled and how individuals’ internalize these processes to then control themselves. In Fotopolou & Riordan’s auto-ethnographic approach, the study finds that Fitbit data provides an account of how much to move, sleep, and eat; as a result, the collection of Fitbit data is a micro-practice with a pedagogical dimension that concerns normalizing and disciplining bodies. Furthermore, in Ruckenstein’s study of participants who are given activity and heart rate monitors, individuals discussed a desire to use self-tracking devices as an external motivator. The researcher states, “For many participants the tracking technology became a daily companion during the research process, a silent persuader that made sure the daily workout was not missed, and that one walked to work instead of driving or taking public transport.”

Our research will expand on these findings and examine the experiences of individuals who have not used fitness tracking devices before, similar to Ruckenstein’s empirical study. However, our study will differ from existing literature through using college-aged participants.

Through the use of self-tracking devices, bodies and lives become more transparent, and thus they can be acknowledged and acted upon. In addition, digital health technologies act as mediators that open aspects of one’s everyday life to scrutiny and intervention, connecting with the theme of self-optimization. Mobile health technologies support a reflexive, self-monitoring awareness of the body. Such technologies bring the body to the forefront as its functions, movements, and habits are constantly monitored, and users are made continually aware of their body via feedback. With the aid of digital technology, Ruckenstein argues that “it is not enough to have a more transparent view of oneself, one needs to respond to that knowledge and raise one’s goals, thereby framing the ‘natural’ body as incomplete.” Therefore, this research examines how data are both produced and consumed by fitness tracking device users, and how these processes result in a form of internal control over the body.

Socialization around health and fitness

Our research also examines how social networks and socialization processes can influence one’s relationship with health and fitness as well as one’s relationship with self-tracking. Scholars argue that socialization processes, social networks and social structures play a role in self-objectification. The dual pathway model suggests that perceived pressure to be thin promotes internalization of the thin ideal. Furthermore, the tripartite influence model hypothesizes that family, peer, and media pressures to be thin results in the internalization of cultural standards of attractiveness and heightened appearance comparison. Research suggests that children and adolescents learn from their families and friends that they should be thin, and that being overweight is not desirable. Within the framework of self-objectification theory, these two models explain how individuals internalize messages that promote attractiveness, and how socialization processes and social pressures to be thin result in body surveillance and self-objectification.

Furthermore, scholars suggest that both women and men experience sexual objectification. In samples of mostly white men, internalization of dominant cultural standards of attractiveness relates to self-objectification, body dissatisfaction, and drive for muscularity. However, research suggests that heterosexual women and minority men report higher levels of self-objectification. Our research will incorporate these ideas in the context of self-tracking devices. Our research will examine whether socialization processes and social networks, such as one’s family and friends, result in the drive to self-track and monitor the body.

Emotions

Our research examines emotions in the context of self-tracking: engaging with a “digital interface” that represents the body and its activity can give rise to certain feelings, both negative and positive. Arlie Hochschild’s work builds on the idea that emotions are always a part of social life, and we must manage our emotions to react the way society expects us to. Hochschild argues, “sometimes we try to stir up a feeling we wish we had, and at other times we try to block or weaken a feeling we wish we did not have.” Hochschild’s ideas of emotion work are relevant to my research as neoliberal values reinforce the idea that people should feel certain ways, such as guilty or shameful, when they are unable to maintain the image of the “healthy” and “good” citizen.

Hochschild’s concept of feeling rules addresses how socially shared norms often obligate individuals to feel a certain way in given situations. Hochschild argues that we recognize a feeling rule “by inspecting
how we assess our feelings, how other people assess our emotional display, and by sanctions issuing ourselves from them.” Furthermore, Hochschild's idea of the “pinch” refers to the discrepancy between “what do I feel” and “what I should feel.” Therefore, we apply Hochschild's idea of socially created emotions to self-tracking devices: we hypothesize that participants will have certain feeling rules based on their generated data, and will conform to social scripts to feel certain ways about that data. For example, because social institutions frame health as a moral obligation in which individuals must take up, socially shared norms reinforce the belief those who fail to address and maintain health in a socially acceptable way, or have an overweight body, are irresponsible and lazy. In the same sense, users of self-tracking devices may feel guilty or shameful when unable to meet certain goals due to the socially shared norm that individuals should never be content with the way they look and should try to better themselves and their bodies. Finally, in the context of self-tracking, participants may feel the pinch when engaging with devices. While society encourages individuals to self-track and quantify health as a “good” behavior, users may not actually feel this way about putting health in numbers, even though these devices normalize that they should feel this way.

**Quantitative Methods**

Our research uses a multi method approach in order to answer our research questions. Our quantitative work captures the larger patterns regarding health and fitness as well as fitness tracking use among students at this liberal arts college, and our qualitative work explains why certain patterns exist. In order to answer our first research question, who are fitness tracking users and do their health behaviors differ from non-users, we use survey data to understand characteristics of device users; how these devices are used; and the relationship between user status and health behaviors. We then use qualitative work to understand why users may have different health behaviors than non-users, such a heightened sense of responsibility to engage in pursuits of fit living.

Our second research question examines whether family and friends influence the adoption of fitness trackers, and also whether family and friends have an impact on one’s relationship with health and fitness as well as self-tracking. Our quantitative work captures the main factors that influence users to begin using tracking devices, as well as importance of health and fitness for family and friends. Simultaneously, our qualitative work adds to the data by examining socialization around health and also how family and friends can influence relationships with self-tracking.

Our third research question examines the implications of self-tracking devices, such as how they influence emotions. Because emotions are difficult to examine through survey data, our qualitative work examines the experiences of users, especially new users, because their experiences are so fresh. Our research also looks at agency in the context of self-tracking devices, which we address qualitatively to understand how users make decisions to disengage from their device.

Our survey includes questions that measure demographic characteristics, such as race, gender, and socioeconomic status. In addition, the survey uncovers descriptive characteristics of tracker use (such as non user, current user, and previous user), reasons for adopting a self-tracking device, decision to use a device, time spent using a device, and activity levels before and after using a device. Therefore, our survey uncovers the prevalence of self-tracking devices among participants, and how the use of these devices differs across race, gender, and socioeconomic status. All respondents in the survey, both users, previous users, and non users, indicate health behaviors, such as amount of exercise, importance of health and fitness, and health/fitness related goals, such as whether individual is trying to lose weight, maintain weight, build muscle, gain weight, or not trying to do anything about their weight. Therefore, we compare variables such as value of fitness, exercise frequency, and fitness goals across non-users, previous users, and current users of self-tracking devices.

In addition, our survey contains questions that measure family and peer socialization, such as whether an individual’s family members and friends use self-tracking devices, and the importance of health and fitness to participants’ family members and peers. Therefore, we also examine the relationship between family and peers’ importance of health and fitness across current users, previous users, and non-users. We hypothesize that individuals who currently use self-tracking devices will report that their families are
In order to conduct the survey, a systematic random sample was drawn from a list of the college student body. Every sixth person out of the list of around 3,000 students was randomly selected in order to gather a sample of 500 students. In total, we received 200 responses from the student body.

Our sample contains college students that predominantly come from upper class backgrounds, are female, are white, and played sports in high school. The survey contains a majority of females (63.7 percent) and a minority of males (36.3 percent). Table 1 shows that an overwhelming majority of participants are white (77.9 percent), 7.9 percent of respondents are Asian or Pacific Islander, 4.7 percent of respondents are Black/African American, 4.7 percent are Hispanic or Latino, and 4.7 percent identify with multiple groups. Furthermore, a majority of students state they come from high socioeconomic backgrounds, as 66.3 percent indicate that in terms of economic resources, they grew up with “more than enough.” Finally, an overwhelming majority of students played sports in high school (90.8 percent).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Number of cases</th>
<th>Percent/mean</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>69</td>
<td>36.3%</td>
</tr>
<tr>
<td>Female</td>
<td>121</td>
<td>63.7%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>190</td>
<td>100.0%</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>148</td>
<td>77.9%</td>
</tr>
<tr>
<td>African American</td>
<td>9</td>
<td>4.7%</td>
</tr>
<tr>
<td>Asian</td>
<td>15</td>
<td>7.9%</td>
</tr>
<tr>
<td>Hispanic or Latinx</td>
<td>9</td>
<td>4.7%</td>
</tr>
<tr>
<td>Identify with multiple groups</td>
<td>9</td>
<td>4.7%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>190</td>
<td>100.0%</td>
</tr>
<tr>
<td><strong>Socioeconomic status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>More than enough</td>
<td>126</td>
<td>66.30%</td>
</tr>
<tr>
<td>Enough and less than enough</td>
<td>64</td>
<td>33.7%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>190</td>
<td>100.0%</td>
</tr>
<tr>
<td><strong>Sports in high school</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>167</td>
<td>90.8%</td>
</tr>
<tr>
<td>No</td>
<td>17</td>
<td>9.2%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>184</td>
<td>100%</td>
</tr>
</tbody>
</table>

Source: self-tracking device user survey
Table 2: Quantitative study measures

<table>
<thead>
<tr>
<th>Variable</th>
<th>Independent Variable/ Dependent Variable/ Control</th>
<th>Conceptualization</th>
<th>Operationalization/Recode in SPSS</th>
<th>Level of measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>User status</td>
<td>IV</td>
<td>Device use is conceptualized as whether an individual currently uses a self-tracking device, has previously used a self-tracking device, or does not use a self-tracking device.</td>
<td>Participants are asked in Q1 if they “currently use a wearable fitness tracking device that monitors steps taken per day, calories burned, sleep patterns, or heart rate.” This question is coded as (1=yes, 0=no) to indicate whether the participant uses a self-tracking device. Participants are asked in Q2 if they “have ever used a self-tracking device in the past but no longer do so”, and answers are coded as (1=yes, 0=no). In order to create the variable self-tracking device users, I computed a new variable so that 1=1 (from Q1 to capture current users), 1=2 (from Q2 to capture previous users), and 0=3 (from Q2 to capture non-users).</td>
<td>Nominal</td>
</tr>
<tr>
<td>Frequency of device use</td>
<td>IV</td>
<td>Frequency of device use is conceptualized as how often an individual uses their self-tracking device on a day-to-day basis.</td>
<td>Participants are asked, “which of the following best describes how often you use your fitness self-tracking device?” Responses are coded (0=never, 1=once a month, 2=every two weeks, 3=once a week, 4=3x a week, 5=5x a week, 6=every day, 98=other, 99=prefer not to disclose). In SPSS I recoded never, once a month, every two weeks, once a week, 3x a week, 5x a week=0, and everyday=1. Therefore, I have intermittent users, or individuals who use their devices irregularly, and everyday users, individuals who use their devices consistently.</td>
<td>Ordinal</td>
</tr>
<tr>
<td>Variable</td>
<td>IV</td>
<td>Description</td>
<td>Scale</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-------------------</td>
<td></td>
</tr>
<tr>
<td>Time spent using device</td>
<td></td>
<td>Participants are asked “how long have you been using the device” and indicate (1= 0 to 3 months, 2= 3 to 6 months, 3= 7 months to a year, 4= 1-2 years, 5=over 2 years, 99=prefer not to disclose). I recoded 1, 2=0 to measure new users, or individuals who have been using the device for less than 6 months, and 3-5 to measure experienced users, or individuals who have been using the device for a significant amount of time, over 7 months.</td>
<td>Ordinal</td>
<td></td>
</tr>
<tr>
<td>Importance of device functions</td>
<td></td>
<td>Participants are asked “in what ways do you use your tracking device, please rank the importance of the following: track weight loss, track sleep patterns, track exercise, log food intake, track heart rate, reminders to move, compete with family members and friends.” Responses are coded (5=extremely important, 4=very important, 3=moderately important, 2=slightly important, 1=not at all important). Due to small response categories, I recoded 5,4=1 extremely and very important, and 3, 2, 1=0 moderately, slightly, and not at all important.</td>
<td>Likert scale</td>
<td></td>
</tr>
<tr>
<td>Influential factors in decision to use device</td>
<td></td>
<td>Participants are asked, “what people or factors influenced your decision to begin using your device? Please rank the importance of the following: family, friends, social media, advertisements, doctor/healthcare provider, and partner.” Responses are coded (7=strongly agree, 6=agree, 5=somewhat agree, 4=neither agree nor disagree, 3=somewhat disagree, 2=disagree, 1=strongly disagree). Due to small response categories, I recoded 7, 6, 5=0 as agree, 4= neither agree nor disagree, and 3, 2, 1=1 as disagree.</td>
<td>Likert scale</td>
<td></td>
</tr>
<tr>
<td>Variable</td>
<td>Type</td>
<td>Description</td>
<td>Details</td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------</td>
<td>------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Activity levels before device use</td>
<td>IV</td>
<td>This variable is conceptualized as how one’s activity levels changed before using a device and after in order to measure whether a fitness tracker results in an increase in activity levels.</td>
<td>Participants are asked, “before using a tracking device, would you say that your activity levels were…” Responses are coded (7=much higher, 6=moderately higher, 5=slightly higher, 4= about the same, 3=slightly lower, 2=moderately lower, 1=much lower). I recoded 7, 6, 5, 4=0 as about the same. The responses for much higher, moderately higher, and slightly higher were extremely low, either 0 or 1, and there were a significant amount of responses for about the same, so this response most accurately reflects the data trends. I recoded 3, 2, 1 as slightly lower, because there were a significant amount of responses for slightly lower, but very few (0 or 1) for much lower and moderately lower. Therefore, slightly lower best reflects the data patterns.</td>
<td></td>
</tr>
<tr>
<td>Family/peer self-tracking device use</td>
<td>IV</td>
<td>This variable measures if current and previous users friends and family members use self-tracking devices.</td>
<td>Participants are asked if their family members and if their close friends use fitness tracking devices. Responses are coded (1=yes, 0=no).</td>
<td></td>
</tr>
<tr>
<td>Importance of health and fitness</td>
<td>DV</td>
<td>This variable is conceptualized as how important health and fitness is to participants, as well as participants’ friends and family members in order to measure socialization and peer influence.</td>
<td>Participants are asked how important is health and fitness to you, how important is health and fitness to your family, and how important is health and fitness to your friends. Responses are coded (5=extremely important, 4=very important, 3=moderately important, 2=slightly important, 1=not at all important). I recoded 5,4=0 as very and extremely important, and 3, 2, 1 as not at all, slightly, and moderately important due to small response categories.</td>
<td></td>
</tr>
<tr>
<td>Exercise frequency</td>
<td>DV</td>
<td>This variable is conceptualized as how often participants workout, meaning go to the gym, attend workout classes, workout at home, or workout however they define working out.</td>
<td>This variable is operationalized as a subjective measure of frequency of working out. Responses are coded (0=never, 1=rarely, 2=sometimes, 3=often, 4= always). Due to small response categories, I recoded 0, 1, 2= 1 as never, rarely, and sometimes and 3,4 as often and always.</td>
<td></td>
</tr>
<tr>
<td>Variable</td>
<td>Type</td>
<td>Description</td>
<td>Example</td>
<td></td>
</tr>
<tr>
<td>------------------------------</td>
<td>---------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Body goals</td>
<td>DV</td>
<td>This variable is conceptualized as the ways in which participants are engaging in projects of body optimization, or if they are not trying to at all, such as build muscle, lose weight, or gain weight.</td>
<td>Participants are asked, “which of the following currently describes your current fitness goals?” Responses are coded (1=lose weight, 2=gain weight, 3=stay the same weight, 4= build muscle, 5= not trying to do anything about weight). I recoded 1=1, 2,3=2, 4=3, 5=4 because the response category was so small for gain weight that I combined this category with stay the same weight.</td>
<td></td>
</tr>
<tr>
<td>Self-objectification</td>
<td>DV</td>
<td>This variable is conceptualized as one's focus on outward appearance, or how people view themselves as objects rather than human beings.</td>
<td>Participants respond to the following statements that indicate whether one is focused on self-scrutinization. The statements are: during the day, I think about how I look many times; I rarely compare how I look with how other people look; I think more about how my body feels than how my body looks; I often worry about whether the clothes I’m wearing make me look good. Responses are coded as (1=strongly disagree, 2=disagree, 3=somewhat disagree, 4=somewhat agree, 5=agree, 6= strongly agree). Due to small response categories, I recoded 1,2,3=0 as disagree, and 4,5,6=1 as agree.</td>
<td></td>
</tr>
<tr>
<td>Sports in high school</td>
<td>Control</td>
<td>This variable is conceptualized as whether participants played sports in high school.</td>
<td>Participants are asked if they played sports in high school. Responses are coded 1=yes, 0=no.</td>
<td></td>
</tr>
<tr>
<td>D1 athletics</td>
<td>Control</td>
<td>This variable is conceptualized as whether participants play a D1 sport at Colgate.</td>
<td>Participants are asked if they are currently on a D1 athletic team at Colgate. Responses are coded 1=yes, 0=no.</td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Control</td>
<td>This variable is conceptualized as what ethnicity participants best identify with.</td>
<td>Participants are asked to identify their ethnicity. Responses are coded as (1=African American, 2=White, 3=Asian or Pacific Islander, 4=Hispanic or Latinx, 5= Native American, 6= Identify with multiple groups). Due to small response categories, for my bivariate analysis I created two groups, white and people of color to have enough responses to run a crosstab. Therefore, I computed a new variable where 2 (white)=1 and 1,3,4,5,6=2 as people of color.</td>
<td></td>
</tr>
<tr>
<td>Variable</td>
<td>Role</td>
<td>Description</td>
<td>Methodology</td>
<td>Level</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>---------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>Gender</td>
<td>Moderating</td>
<td>This variable is used to test whether the relationship between IV and DVs differ by gender</td>
<td>Participants are asked to indicate which gender they best identify with. Responses are coded as (1=man, 2=woman, 3=transgender, 4=non binary/third gender, 98=other). I recoded 1=1 and 2=2 and there were no responses in the other gender categories.</td>
<td>Nominal</td>
</tr>
<tr>
<td>Socioeconomic status</td>
<td>Control</td>
<td>This variable is conceptualized as participants’ resources they had access to growing up, specifically financial resources.</td>
<td>Participants are asked, “in terms of economic resources, would you say that you grew up with…” Responses are coded as (1=more than enough, 2=enough, 3=less than enough). I recoded 1=1 as more than enough and 2,3=2 as enough and less than enough because of small response categories.</td>
<td>Nominal</td>
</tr>
</tbody>
</table>
**Independent variables**

As shown in Table 2, our independent variables are user status (current, previous, and non-users), frequency of device use, time spent using the device, importance of device functions, and influential factors in decision to use a device. Our independent variables allow us to examine who are the users, how do they use their device, and what factors motivated them to begin using a device. We first separate the sample into three groups: current users, previous users, and non-users. We then use univariate analysis in SPSS to find out from current users how often they use their self-tracking device and how long they have been using the device. Next, we have both current users and previous users indicate which functions of the device are most important to them. Finally, respondents indicate which people or factors most influenced their decision to begin using a tracking device.

**Dependent variables**

As shown in Table 2, our dependent variables are health behaviors as well as family and friend use of self-tracking devices. The health behaviors studied are importance of health and fitness, exercise frequency, and body optimization goals. We use bivariate analysis in SPSS to run health behaviors by the independent variable, user status (current, previous, and non-users). Participants also indicate importance of health and fitness for their friends and families. Therefore, we use bivariate analysis in SPSS to run importance of health and fitness to friends and family by the independent variable user status. We also use bivariate analysis in SPSS to run family and friend use of self-tracking devices by user status in order to examine whether current users have more friends and family members with self-trackers than non-users.

**Qualitative Methods**

While the survey helped us identify larger patterns of health tracking and behaviors in our sample, the main findings from our research come from our qualitative in-depth interviews which provide information on lived experiences and insights about how sample members experience, think about, and grapple with health and fitness technology. In addition, we employed a longitudinal examination of college students’ experience before and after using the device to uncover how these devices influence health behaviors, emotions, and agency among new users.

Our first research question asks whether users and non-users have different health behaviors. Our qualitative work aims to clarify why users and non-users may have different health behaviors. Specifically, we set out to determine whether tracking devices heighten responsibility for health and fitness, as well as encourage users to control themselves internally. Furthermore, our second research question addresses whether family and friends influence adoption of a fitness tracker and influence self-perceived value of health and fitness. Therefore, our qualitative work examines why socialization might influence both health and fitness as well as self-tracking. We ask participants how they viewed health and fitness growing up, as well as in their peer environments at the elite liberal arts college, to get a better understanding of socialization processes. Our third research question investigates the datafied self: how individuals experience fitness tracking technologies and the emotions related to these experiences. We address this question qualitatively and focus on how new users grapple with the technology for the first time. Finally, we focus on agency in our qualitative work as well in order to understand how individuals make independent decisions in the context of self-tracking.

Our interview sample is divided into four different groups in order to represent a wide range of perspectives. Table 3 shows the four groups: new users, previous users, non-users, and current users. New users received a Fitbit for a period of three weeks in order to examine their experience using the device for the first time. New users are most likely to be hyper-aware of their response to fitness trackers as this is a new experience, compared to current users, who have been using the technology for at least three months. Previous users, or college students that formerly used a self-tracking device but no longer do so, were recruited because they have chosen to stop fitness tracking in their lives. Finally, non-users were recruited in order to examine whether their experiences of health and fitness were different from users. Our qualitative work contains data from sixteen total participants: five new users, four non-users, three previous users, and four current users. The non-user and current user groups
each contain two males and two females, while the new user group contains three females and two males, and the previous user group contains one female and two males.

Table 3: Qualitative Groups

<table>
<thead>
<tr>
<th>Group</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>New users</td>
<td>Individuals who are given Fitbits for a period of three weeks and have never used a self-tracking device before</td>
</tr>
<tr>
<td>Experienced users</td>
<td>Individuals who have been using self-tracking devices for an extended period of time (over six months)</td>
</tr>
<tr>
<td>Previous users</td>
<td>Individuals who used self-tracking devices in the past but no longer do so</td>
</tr>
<tr>
<td>Non-users</td>
<td>Individuals who have never used a self-tracking device</td>
</tr>
</tbody>
</table>

Additionally, new users were asked to fill out a health diary so they would be able to easily recall their experiences over the past three weeks and reflect on their feelings in relation to the technology. Over the course of the three weeks, participants were prompted to fill out health diary questions every Monday, Wednesday, and Friday. These questions were: How do you feel today? How did your device make you feel today? How did you use your device today? How would you define health today? We then coded these diaries for patterns in the data.

We first piloted this study with a female non-user who used the Fitbit for three weeks and filled out questions in a health diary in order to ensure the process would run smoothly. After seeing that using the Fitbit did not seem to pose any harm, we then conducted the study with two men and one woman, each of whom were given a Fitbit for a period of three weeks. Ultimately, we aimed to have equal numbers in each group, but we ended up with an additional new user and lacked another female previous user.

Table 4: Qualitative interview participants

<table>
<thead>
<tr>
<th>Name*</th>
<th>Tracker use</th>
<th>Gender</th>
<th>Race</th>
<th>Socio-economic status</th>
</tr>
</thead>
</table>

| Amy   | Experienced user (Apple Watch) | Female | White              | More than enough      |
| Mary  | Experienced user (Apple Watch) | Female | White              | More than enough      |
| Carly | Previous user (Fitbit)        | Female | White              | More than enough      |
| Jamie | Previous user (Fitbit)        | Female | White              | More than enough      |
| Emma  | New user (Fitbit)             | Female | White              | More than enough      |
| Sophie| New user (Fitbit)             | Female | White              | More than enough      |
| Charlotte | New user (Apple Watch) | Female | White              | Enough                |
| Ellie | Non-user                   | Female | White              | More than enough      |
| Rachel| Non-user                   | Female | White              | More than enough      |
| Ian   | Experienced user (Fitbit)     | Male   | Hispanic/Latino    | More than enough      |
| Andrew| Experienced user (Apple Watch) | Male  | African American   | Enough                |
| Paul  | Previous user (heart-rate monitor) | Male  | White              | More than enough      |
| Mark  | Non-user                   | Male   | White              | More than enough      |
| Robert| Non-user                   | Male   | White              | Enough                |
| Thomas| New user (Fitbit)            | Male   | White              | More than enough      |
| Chris | New user (Fitbit)            | Male   | White              | More than enough      |

* Names are changed to pseudonyms

Participants used either Fitbits or Apple Watch-
es as wearable tracking devices. Fitbit offers a range of wearable devices that can be attached to the wrist or clipped onto clothing. Daily use of Fitbit involves wearing the device throughout the day while it monitors steps walked, floors climbed, and calories burned. Users can access their data through their personal dashboard, which can be accessed through the website or on the mobile app. The device has users input their height, weight, and age to determine a specific number of calories to be burned each day. It then sets this calorie threshold and walking a minimum of 10,000 steps per day as personal weight loss goals. The Fitbit sets a minimum of 10,000 steps per day, which equals about 5 miles as a way for the user to create a calorie deficit in line with their caloric intake to lose weight. The Fitbit motivates the user to walk and burn more calories because if you burn more calories than you consume, this deficit results in weight loss. Apple Watches have a broader scope of capabilities, including texting, calling, checking email or using social media; however, they are also marketed for health tracking. The Apple Watch Series 4 is described as “fundamentally re-designed and re-engineered to help you be even more active, healthy, and connected.” Therefore, research participants used both brands of fitness tracking devices.

The participants in the new user group were given the Fitbit Flex throughout the longitudinal examination. Since these devices were purchased for the study, we had to use a basic fitness tracker band without the touchscreen capabilities of the Apple Watch. The Fitbit Flex does not have an interactive screen for the user to receive notifications on throughout the day, and the user must open the app on their phone to see their data and progress. Although it would have been preferable to compare participants with the exact same fitness tracking device, for the experienced user and previous user group it would have been too difficult to do so with a small sample size. It is important to note that one student we interviewed from the survey indicated she was a non-user; however, by the time she was interviewed she had received an Apple Watch over winter break. Therefore, we categorized this student in the new user group as well.

Results

Health Behaviors

Our first research question aims to understand who are tracking device users and what are the differences in health behaviors for users and non-users. In addition, our first research question asks why health behaviors are different for users and non-users. We argue that both a heightened sense of responsibility for health and a hyper-awareness of the body among users are correlated to different health behaviors.

| Table 5. Description statistics of self-tracker users |
|-------------------------|-------------|----------------|
| **Variable** | **Number of Cases** | **Percent/ Mean** |
| **User** | | |
| Current user | 59.00 | 28.9% |
| Previous user | 40.00 | 19.6% |
| Non-user | 105.00 | 51.5% |
| **Total** | 204.00 | 100% |

Source: self-tracking device user survey

First, we evaluate the prevalence of wearable fitness tracking devices on this campus. Table 5 shows that in a sample of predominantly white individuals from high socioeconomic status that attend an elite college, almost half of participants (49%) currently or have previously used a wearable fitness tracker. Twenty-nine percent are current users and 20 percent are previous users. Fifty-two percent of individuals have never used a fitness self-tracking device.
Next, in order to understand who users are, we look at the prevalence of self-trackers across demographic characteristics such as gender, race, and socioeconomic status. Table 6a shows that females (73.2%) are much more likely than males (26.8%) to currently use wearable fitness trackers (p-value>0.05). This suggests that females are more likely to be drawn to quantifying their health in numbers compared to males. Table 6b shows that white people (80.4%) are much more likely to use fitness tracking devices compared to people of color (19.6%) (p-value>0.05). Finally, Table 6c shows that individuals of high socioeconomic status (75%) are much more likely to used wearable fitness trackers (p-value>0.05).

In order to understand who users are, respondents indicate which functions of the device are most important to them. In addition, these findings present initial suggestions towards health behaviors for users, as the functions of these devices are to monitor health and fitness. Table 8 shows that tracking exercise (87%) and tracking heart rate (45.6%) are rated as the most important. The importance of tracking exercise is unsurprising due to the pressure students place on themselves to maintain their fitness and workout routines, which will be explained further in our qualitative work. These two features are followed by tracking sleep patterns (33.3%), reminders to move (23%), compete with family members and friends (11%), track weight loss (9%), and finally log food intake (7%).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Number of Cases</th>
<th>Percent/ Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Track weight loss</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extremely and very important</td>
<td>5.60</td>
<td>8.8%</td>
</tr>
<tr>
<td>Not at all, slightly, and moderately important</td>
<td>52.90</td>
<td>91.2%</td>
</tr>
<tr>
<td>Total</td>
<td>57.50</td>
<td>100.0%</td>
</tr>
<tr>
<td>Track sleep patterns</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extremely and very important</td>
<td>19.90</td>
<td>33.3%</td>
</tr>
<tr>
<td>Not at all, slightly, and moderately important</td>
<td>38.90</td>
<td>66.7%</td>
</tr>
<tr>
<td>Total</td>
<td>57.80</td>
<td>100.0%</td>
</tr>
<tr>
<td>Track exercise</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extremely and very important</td>
<td>50.90</td>
<td>87.7%</td>
</tr>
<tr>
<td>Not at all, slightly, and moderately important</td>
<td>7.60</td>
<td>12.3%</td>
</tr>
<tr>
<td>Total</td>
<td>57.50</td>
<td>100.0%</td>
</tr>
<tr>
<td>Log food intake</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extremely and very important</td>
<td>4.60</td>
<td>7.0%</td>
</tr>
<tr>
<td>Not at all, slightly, and moderately important</td>
<td>53.90</td>
<td>93.9%</td>
</tr>
<tr>
<td>Total</td>
<td>57.50</td>
<td>100.0%</td>
</tr>
<tr>
<td>Track heartbeat</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extremely and very important</td>
<td>26.90</td>
<td>45.6%</td>
</tr>
<tr>
<td>Not at all, slightly, and moderately important</td>
<td>31.90</td>
<td>54.4%</td>
</tr>
<tr>
<td>Total</td>
<td>57.80</td>
<td>100.0%</td>
</tr>
<tr>
<td>Reminders to move</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extremely and very important</td>
<td>13.90</td>
<td>22.8%</td>
</tr>
<tr>
<td>Not at all, slightly, and moderately important</td>
<td>44.90</td>
<td>77.2%</td>
</tr>
<tr>
<td>Total</td>
<td>57.80</td>
<td>100.0%</td>
</tr>
<tr>
<td>Compete with family members and friends</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extremely and very important</td>
<td>6.60</td>
<td>10.5%</td>
</tr>
<tr>
<td>Not at all and moderately important</td>
<td>51.90</td>
<td>89.5%</td>
</tr>
<tr>
<td>Total</td>
<td>57.50</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Source: self-tracking device user survey
Next, we look at the value of health and fitness among all respondents in order to understand health behaviors of the entire sample. Table 9 shows that among all participants, health and fitness is highly valued. Over half of all respondents (70.8%), regardless of tracker use, indicate that health and fitness is very or extremely important. In addition, over half of the sample (67.5%) indicate that they work out often or always. Finally, a majority of participants in the sample are engaging in projects of body transformation. 72.7 percent of respondents are trying to change their bodies in some way, either by losing weight (35%) or building muscle or strength (37.7%). Therefore, a consistent fitness routine and projects of self-optimization are highly normalized on campus.

Next, we examine whether these health behaviors are different for users and non-users. Table 10 shows that a higher percentage of current self-tracking device users (83%) indicate that health and fitness is extremely or very important compared to non-users (62%) (p-value<0.05). In addition, Table 10 shows that current self-tracker users (77%) exercise often and always more than non-users (61%) (p-value>0.05). Current self-tracker users may exercise more than non-users due to being highly invested in health and fitness, and because the device is a constant reminder of the need to workout and meet certain goals. There were no significant differences in fitness goals between users and non-users.

While our quantitative work examines health behaviors for users and non-users, our qualitative work attempts to get at the narratives behind why health behaviors differ between the two groups. First, it is important to unpack why most participants in our qualitative work, regardless of user status, feel highly responsible for health. These findings are consistent with our quantitative results among the entire sample, which show that a majority of participants exercise frequently and believe that health and fitness is very important. However, the ways in which current and new users discuss their tracking devices, and that they feel a heightened responsibility for health after using the device, lead us to believe that the device is related to different health behaviors. In addition, our quantitative work supports the idea that health behaviors differ for users and non-users, as users exercise more often than non-users and place a higher value on health and fitness compared to non-users. Therefore, we compare the attitudes surrounding health as a responsibility across all groups. In the last part of this section, we
focus on how self-tracking device users have a heightened sense of responsibility and are hyper-aware of the body. We argue that these mechanisms of social control are related to the different health behaviors between users and non-users. These mechanisms of social control manifest in the ways these devices act as a constant reminder to be the neoliberal “healthy” and “good” citizen by achieving certain quantitative metrics, and failing to do so can reinforce or create negative beliefs about the self.

Health as Responsibility

All four groups of our participants discuss the pressure they place on themselves to keep up with their fitness routines. Respondents indicate that health and fitness is highly valued, and that they exercise frequently. Living together on a residential Division I sports campus, all experience pressure to engage in fit living, to “take up” health and fitness in pursuit of the regulated, responsible, and productive citizen. For example, Robert, a male non-user explains that he would feel upset if he had to miss a workout, explaining “It’s like, oh, you know, if I don’t work out I’m all of a sudden going to get fat and I have to remind myself, no, you’re not going to get fat.” With rising concerns over the “obesity epidemic,” moral imperatives exist which compel individuals to be healthy and stigmatize the overweight body as “unhealthy.” This form of bio-power is apparent through Robert’s internal dialogue: he feels as though he is doing something wrong by not working out, and that there will be consequences such as gaining weight. Therefore, rising societal concerns over obesity and “getting fat” act as a form of social control on a macro-level, which then pushes college students to control themselves and fear not keeping up with their health and fitness routines.

In addition, new users, experienced users, previous users, and non-users all discuss how health and fitness was something they felt they needed to “take up” recently when they left home and began a life in college. Ian, an experienced Fitbit user, says, “Once I got to college I was away from home so I wanted to be able to take care of myself, not just like freshman fifteen and just ruin it.” Ian discusses how he felt the need to start going to the gym in college so he didn’t gain the “freshman fifteen,” a common phrase for gaining weight the first year in college. The act of taking up health and fitness as one leaves home reflects how it is up to young adults to engage in projects of self-optimization, be “healthy” citizens, and fear gaining weight. Mary, an experienced self-tracker user, states:

Once I stopped playing sports and my practices weren’t mapped out for me, like freshman year of college I needed to start actually working out on my own which was really weird…I got my Fitbit then so it was like it kind of, I felt responsible, and then in order to kind of keep myself in check I got the Fitbit and then the Apple Watch, so kind of just to map out how I was doing because I didn’t have like an athletic performance to like tell me if I’m doing well.

Mary’s discourse reflects the idea of healthism, or how college students are held morally responsible for their own health, and are expected to acquire knowledge and adopt self-awareness of how to maintain health. For Mary, she feels responsible to ensure that she keeps herself “in check” and in order to do so uses a wearable fitness tracking device. In addition, Mary feels the need to measure how well she does in regards to her workout regime, since she no longer has organized sports to measure her athletic performance. Therefore, the wearable fitness tracker acts as a mechanism to measure and evaluate the self, reflecting the idea that individuals should push themselves harder and be unsatisfied with the way they are. As a result, college students use self-tracking devices to monitor themselves and improve their lives in some way.

What is most interesting regarding personal responsibility for health is how current and new users of fitness self-tracking devices discuss a heightened sense of responsibility after using the device. In our quantitative work, a higher percentage of current users (82.5%) indicate that health and fitness is extremely or very important compared to non-users (61.8%). This finding can be explained by the fact that fitness trackers make one think about health and fitness more, and promote a hyper-awareness of the body. Sophie, a new Fitbit user, discusses how she felt more responsible after using the device for a period of three weeks. When I asked her if she felt a higher sense of responsibility after using the device, she says, “I think so… I think maybe using the device made me more like hyperactive about [health and fitness]. Like maybe think about it more.” Therefore, although Sophie says that she cared
about health both before and after using the device, it made her feel more responsible because she was thinking about health and fitness more. When I asked what she means by “think” about health more, she explains:

*It was like 24/7, even if I was at the library, like most of the time I’ll know during the day, I’ll be like, okay, at this time I’ll go to the gym. But like if [the Fitbit] is there I’m like, oh, I could go to the gym right now, I know I should work out right now. Or if it’s telling you to move, like I should be doing something, not just sitting here.*

Therefore, using the self-tracking device invites a heightened awareness of health and fitness for Sophie, and looking at her device resulted in constant thoughts regarding fitness rituals. Sophie’s experience reflects the theory of biopedagogy, and how wearable fitness self-tracking devices train and teach our bodies how to live.53 In this context, the wearable fitness tracker trains Sophie to feel like she should be “doing something” and be productive, reflecting the neoliberal approach to being the “healthy” and “good” citizen. Therefore, fitness tracking devices add another layer of social control that encourages users to self-policing. Although all participants self-police to some degree, the device acts as a visual representation of the need to control oneself and meet certain goals.

Emma, another female new Fitbit user, also felt that using a fitness-self tracking device made her feel more responsible, and that she should be taking her health and fitness to the next level. When I asked if she felt more responsible at the end of the three weeks post using the device, she says:

*Yeah, definitely. I think that the Fitbit made me more aware of like every single choice that I was making versus a more holistic view...like every little thing makes a difference. So I think there is good and bad in that. I think it motivated me to push myself and make those little changes, but also to the point where like if it’s not convenient for me to do it I was still doing it to take my fitness to the next level.*

Emma discusses the fact that the Fitbit made her feel as though she had to take her fitness to the “next level.” The discourses of new users are extremely important: these participants have their experiences and feelings fresh on their minds as they had just used the device for the first time. Therefore, these participants recognize the stark changes in themselves from before using the device and after. For Emma, the Fitbit heightened responsibility because she was made aware of her choices on the device, versus before using the device she had a more “holistic” view and would not have to zone in on “every little thing.” In addition, Emma discusses how the Fitbit encouraged her to “push herself,” demonstrating how in a neoliberal society, individuals are expected to engage in technologies of self in the pursuit of bettering oneself in a manner indicated by societal institutions. Therefore, this phenomenon reinforces docile bodies as individuals can blindly follow such technology to reach the “next level” and optimize. We see in Emma’s story how self-tracking devices rationalize a framing of the body as deficient and in need of improvement, deepening responsibility for health and fitness.54 While Emma most likely policed herself before using the fitness tracker to some degree, this level of self-policing was heightened when using the device, resulting in a drive to take her fitness to the next level, which suggests an increased exercise frequency and a heightened emphasis on health and fitness when using the device.

**Hyper-awareness of the Body**

New, previous, and experienced users discuss how using a wearable fitness tracker makes them hyper-aware of the body’s functions and movements. We argue that this form of hyper-awareness is related to different health behaviors between users and non-users. The ability to see one’s body’s functions and movements throughout the day increases responsibility for health, fitness, and self-optimization. Amy, an experienced Apple Watch user, states, “I guess [the Apple Watch] made me more aware of what I was doing… the stuff on the machines are never that accurate so I would only know from what I was doing… the stuff on the machines are never that accurate so I would only know from what I was doing at the gym.” For Amy, before she used a self-tracking device, she would see that she burned a certain amount of calories at the gym and that was all she would know about her activity for the day. However, she explains that with her tracking device, she learns a lot more about her body’s movements, saying, “Now I can see like I really didn’t walk today like I didn’t even sit up and it’s like when I’m studying I get all those reminders that I haven’t
walked in an hour or I can see my average heart rate… and like you get your calories not just from your work-out but for the whole day.” Amy’s self-tracking device does not allow her to avoid the fact that she “didn’t walk today,” representing how users are made consistently aware of their body and its movements.

Ruckenstein argues that self-tracking devices make our bodies and lives much more apparent, and therefore they can be better acted upon. The use of a self-tracking device brings the body to the forefront of the mind as its movements and functions are datafied throughout the day, intensifying responsibility for fit living. As a result, users have different health behaviors than non-users because of the pressing reminder on their wrist that they should be working out and productive. The device heightens responsibility for fitness and makes it much more difficult to disengage from self-policing.

Charlotte, a new Apple Watch user, also speaks about her device’s ability to bring activity to the forefront of the mind, which can intensify health and fitness as a responsibility of the individual. She explains that she thinks about her activity more with her device: I definitely am much more conscious on the days where I’m like, for example, my, one of my friends got a Playstation for Christmas for like, God knows why, but she brought it here on Saturday and all of us just sat on that thing for like 10 hours. And it was like I looked at my watch at the end of the day and it was like, you moved like four hours a day out of 10. I was like, oh my God, like I haven’t done anything today.

Charlotte also explains that before having an Apple Watch, she wouldn’t think about not moving enough during the day. The use of self-tracking devices promotes a hyper-awareness of the body and its movements, and therefore, one’s cognitive state becomes much more focused on movement and activity in the context of regular technological feedback. Ruckenstein’s study had similar findings: the researcher discovered that the monitoring equipment made people much more aware of their bodies and how their bodies felt. Therefore, the influence of wearable fitness trackers on college students’ lives becomes evident; users feel they must constantly confront and optimize their activity as seen through a digital interface, adding an additional layer of social control in the context of health and fitness.

Selfhood

Our next research question addresses how individuals are socialized around health and fitness, and how socialization plays a role in self-tracking. In this section, we first quantitatively analyze the importance of health and fitness for participants’ families and friends, and then qualitatively present our results regarding socialization around health and fitness as well as self-tracking.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Current user</th>
<th>Previous user</th>
<th>Non-users</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Female health/fitness</strong></td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td>Not at all, slightly, and not important</td>
<td>21 (77)</td>
<td>11 (34.4)</td>
<td>30 (35.8)</td>
<td>51 (36.2)</td>
</tr>
<tr>
<td>Very and extremely important</td>
<td>16 (63.6)</td>
<td>35 (49.2)</td>
<td>64 (82.8)</td>
<td>125 (63.8)</td>
</tr>
<tr>
<td><strong>Peer health/fitness</strong></td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td>Not at all, slightly, and not important</td>
<td>34 (64)</td>
<td>12 (22)</td>
<td>44 (80)</td>
<td>90 (51.3)</td>
</tr>
<tr>
<td>Very and extremely important</td>
<td>27 (40.4)</td>
<td>24 (40.4)</td>
<td>52 (52)</td>
<td>103 (51.3)</td>
</tr>
</tbody>
</table>

Our quantitative work first measures the importance of health and fitness for participants’ friends and families in order to examine socialization in this context. All participants’ family members and friends place a high value on health and fitness. Table 11 displays that over half of all respondents (63.8%) indicate that health and fitness is very or extremely important to their family (p-value<0.05), and over half of all respondents (51.3%) indicate that health and fitness is very or extremely important to their peers (p-value<0.05). However, there were no significant differences in the value of health and fitness for users’ friends and families compared to non-users. This was surprising because we expected that users of self-tracking devices would have families and friends that value health and fitness more than non-users. Overall, these results highlight the centrality of social learning for all participants, such as how families and friends influence the importance of health and fitness for individuals.
Users are most influenced to adopt fitness tracking by the people closest to them, such as friends and family members, rather than larger macro influences such as advertisements or healthcare providers. Table 13 shows that over half of the respondents in the sample were influenced by family (57%) and friends (55%). Doctors (7%) and partners (9%) influenced respondents at much lower rates than compared to their family members and friends. Furthermore, less than half of respondents were influenced by other factors such as social media (43%) or advertisements (37%). This shows that individuals who use self-tracking devices are highly influenced by people in their closest networks.

Although our survey research shows that individuals who use self-tracking devices are highly influenced by their friends and family, it is limited in its ability to show whether family and friends influence relationships with self-tracking. As mentioned previously, most participants emphasize that their friends and families are highly invested in health and fitness. What is most interesting, however, is we believe that certain kinds of parenting and peer groups can lead to adopting a self-tracking device, and also feeling more obsessive about quantifying health. Without the parental figure to tell children what they should and shouldn't be doing, the fitness tracker takes the place of the helicopter parent and allows individuals to monitor themselves.

Our major finding regarding socialization and self-tracking is that among participants, individuals with higher peer group or family pressures to self-optimize appear to have different relationships with wearable fitness trackers than individuals who do not feel pressure from their friends and family. However, our sample size is too small to generalize these findings. Amy, a female, experienced Apple Watch user, discusses how she had a difficult time with her weight at a young age, how growing up her parents were very fitness-oriented, and how at the liberal arts college her friends can be toxic regarding their eating and workout habits. Amy discusses her peer environment, and how the people she surrounds herself with often make comments that make her feel bad. She explains: “In the spring, last semester a girl was like ‘oh my god I literally was so drunk and drunk ate an English muffin,’ and I’m like I just chose to eat one for breakfast, but for her that was like so bad, and it’s like I just ate one and you’re like what the hell that’s so annoying.” In addition, Amy discusses how the peers in her environment discuss going to the gym as a way to punish themselves for bad health behaviors, such as eating certain unhealthy foods. Therefore, the people in Amy’s environment engage in high levels of self-policing, and feel guilt and shame for engaging in “bad” health behaviors.

Amy also discusses how she shares her activity with her friends on her fitness tracker, and suggests a level competitiveness. She says “I’ll see people will have like such a good high calorie burn… and I might not have burned a lot of calories even though it was still a good workout, so I’m like ugh shoot I should have done something else, so sometimes it like bothers me.” Therefore, Amy is not only policing herself with her own device, but is also policing herself by comparing...
her workouts to her friends’ workouts. A high level of self-policing, combined with a peer environment where this behavior is normalized, may influence relationships with fitness tracking.

The context of Amy’s socialization and peer environment relates to the way she engages with health and fitness and her fitness tracker. Amy uses her Apple Watch as a way to ensure that she stays on track and meets her health and fitness goals. In the interview, it was apparent that her watch was a significant part of her life. When we asked her how she would feel if she wasn’t wearing her watch, she says: “Oh I feel lost! Like there’s sometimes when I’ll go to workout and I’m like oh my god I’m so mad because it’s like yes, this workout counts, it’s so much more than the watch. But it’s like I want it logged on my watch…’and I’m like ugh my goal isn’t going to be met.” Her discourse represents how the device is a form of internal control because she feels the need to have it on all of the time to meet certain quantified and measurable goals. Amy feels the need to control herself internally through self-tracking due to her family’s pressure on fitness and her competitive peer environment. As a result, outside factors and forces can influence the relationship one has with their self-tracking device.

On the other hand, participants with families and peers who take a more relaxed approach toward health and fitness have different relationships with their wearable devices. Charlotte discusses her relaxed peer environment, such as how her friends at college aren’t toxic about their gym-going habits. She describes, “I would say my friends are pretty relaxed about health and fitness. I have a couple who like um, do like IM sports here that they’re pretty into, but like no one is, you know, like I have to go to the gym every day of the week type of thing.” Charlotte’s peer group norms and socialization processes were much different from Amy’s. In addition, when discussing Charlotte's Apple Watch, she explains that she only would mind not having her watch because she wouldn’t be able to tell the time, and that missing out on some fitness data isn’t a big deal to her. She also says that working out is more about feeling and not about quantifying health, and doesn’t seem to care too much about the fitness tracker. When I asked her how she resists having an obsessive mentality about her watch, she states:

_I honestly would say it’s the people I surround myself with. Like I mentioned my friends, if they work out it’s because like they have time to and it’s just like something that they enjoy... I would say it’s definitely because of none of them have the kind of like I have to go to the gym and if I don’t like everything is ruined. I would say it’s also because I’m surrounded by people who are like really calm about their workout routines and like my parents were never like they went to the gym every day or whatever so it was nothing like that. I think it’s because of those people that I’m able to sort of adopt that similar, like pretty calm, pretty chill vibe about going and working out in whatever sense that is._

Charlotte differs from Amy in that she wouldn’t feel lost or blind without her device, and it was apparent in the interview that self-tracking wasn’t a major part of her life. She wasn’t using her device to engage in projects of self-optimization, but more as a way to engage with her data in a positive way as she only works out to feel good about herself. She feels that she doesn’t care about fit living to an extreme level because of the people she surrounds herself with, as her friends also go to the gym because it’s something they enjoy. Different from Amy’s friends, her friends don’t go to the gym as a way to punish themselves for bad health behaviors. Therefore, one's environment and socialization processes may affect the way in which one approaches health and fitness, especially in the context of self-tracking devices.

Grappling with Datafied Selves

Fitness tracking devices can result in feelings of shame when users see a lack of activity on the device. Although both new and experienced users discuss emotions in the context of their wearable fitness tracker, new users had the most heightened emotions because they had just used these devices for the first time, and could easily reflect back on specific experiences where they felt the effects of the tracking device. Chris, a new Fitbit user, discusses his experience using the device for three weeks:

_I remember one day I went out the night before, barely slept, went to an early class, and then came home after class and slept. And when I woke up I noticed that my step count was only at 0.5 miles, and it made me feel like a fat slob._
Chris sees that he has only walked 0.5 miles, which is a visualization of inferiority through the medium of digital data. Therefore, because Chris sees that he is far from his goal (the Fitbit sets the goal to around 5 miles a day or it can be customized based on the user's personal weight loss goals depending on their gender, height and weight), the device illustrates that his results are unsatisfactory, prompting him to confront and transform that lack of data. To Chris, his lack of activity results in feelings of shame, as he feels like a “fat slob.” Chris’s feelings relate to Hochshild’s work around feeling rules, and how people are expected to feel certain ways based on societal scripts. Contributing to Chris’s shame may be neoliberal beliefs that perpetuate the idea that responsible students should be productive and healthy.

Fitness tracking devices also influence emotions in other ways: college students discuss feeling lazy when they see certain results, and this data visualization reinforces that people should be productive. Emma, another new Fitbit user, discusses her engagement with the data on the device. She says,

On the nights where I would get a lot of hours of sleep, like 10 hours or something, I would be like oh my god I’m wasting so much time sleeping. Like that kind of thing made me feel lazy in that way.

Therefore, the use of self-tracking devices and seeing oneself through a digital interface brings out emotions, such as feeling lazy for sleeping too much, that may not be as heightened without seeing the body on an app or device. Therefore, the constant monitoring of the body and its functions through using wearable fitness technology can result in negative emotions.

Engaging with one’s data may result in negative emotions due to internalizing a “techno-gaze.” The “techno-gaze” refers to how tracking devices offer flows of data regarding one’s personal information, which encourage and persuade users to engage with their data. In addition, the techno-gaze seemed to be the most heightened for new users because of their newness to the device; experienced users had been using these devices for so long that this process had become normalized. Emma, a new Fitbit user, discusses how she felt on the days that she didn’t reach her Fitbit goals:

I also think it’s harder to ignore or pretend that you don’t see it. Because looking at the app, it’s a very visible way to see what you’ve done, and days that you miss it’s clearly, like I don’t know it makes me feel more guilty in a way. The way they do the graphs and numbers it’s like I don’t know, in like a weird way it’s kind of like school, so it made me like, it was that extra push to go. And seeing it on the app almost makes it worse because it’s very visible and you can’t like ignore it.

Emma’s Fitbit shows her body’s activity levels in bar graph form, which lays out each day of the week by the number of steps taken per day. If a student doesn’t meet the goals one day, that bar will be lower than the others and won’t show that the goal was completed. The visual representation of data makes it very hard for participants to ignore a lack of activity on a certain day. In addition, Emma not only cannot ignore the fact that she didn’t workout, but she also says that being unable to meet a goal “makes me feel more guilty in a way.”

Emma’s experience reflects the theory of biopedagogy and how these devices train and teach our bodies how to live. Furthermore, these devices push users to change their behavior and adopt a fit lifestyle, creating conditions for further feedback and perpetuating the cycle of interactivity. Therefore, this device acts as a form of social control because it encourages users to confront their activity levels and change their behaviors. This process creates a never-ending cycle of data prosumption (where the user creates the data through activity) and data consumption (where the user looks at their data and reflects on it). In addition, engaging in this never-ending cycle results in feelings of shame and guilt if the user is unable to meet their activity goals. Emma also discusses how seeing the graphs and numbers laid out in this way is “kind of like school.” Due to the competitive nature of this elite campus where students are continually attempting to optimize themselves, their grades, their careers, and their lives, it makes sense that students in this environment would want to optimize the data they see on their wearable fitness tracker, and feel shame if unable to do so.

On the other hand, many participants discuss how seeing themselves through a digital interface, or seeing their activity in digital form, can result in feelings of validation and accomplishment. Mary, a female experienced Apple Watch user, discusses her reason-
ing behind using an Apple Watch: she says the device helps her “assist and assess kind of, like show me okay [Mary] great job you worked out or you reached your goals even if you didn’t go to the gym. You still walked around and moved your body and I know you did enough kind of thing.” Therefore, using an Apple Watch helps Mary assist in achieving her goals, and allows her to assess her day-to-day activity to ensure she is doing enough. Ruckenstein uses the term ‘engagement with data double’ to describe how individuals use self-tracking devices because they are seeking information, guidance, and encouragement in the context of health and fitness; in addition, individuals have a desire to be measured, evaluated, and challenged with data flows.[63]

We add to Ruckenstein’s idea as we believe that college students are not only engaging with their data as a way to be measured and challenged, but also look to their stats as a form of validation that makes them feel accomplished and successful. We found that college students also use the fitness tracker as a way to stay “on track” and achieve success without parents, coaches and the structure they had in high school. Mary’s discourse connects to the idea of concerted cultivation and the need to replace the ‘helicopter parent’ with a device that will help her stay on track. The device validates the fact that she is working out without a coach or a parent to give her a high five at the end of a game or practice. Furthermore, this shows how Mary wants to be measured and evaluated by her watch in order to know if she's doing a “great job.” Without a coach, parent, or sport to tell Mary that she is succeeding, the Apple Watch acts as this form of validation to tell her she’s doing fitness “right.”

Experienced and new users discuss how quantifying health is motivating and results in positive feelings because users can see the payoffs of their hard work in digital form. Chris, a male new Fitbit user says that he feels better when he works out with his device rather than without it. When I asked him why it makes him feel better, he responded: “Because you can see how much you’re doing and it makes me feel better about myself.” He continued to describe how it makes him feel better about himself, saying, “Just like seeing that number go up is nice, kind of like looking at your bank account and seeing your number go up… a little bit of a, kind of like a satisfaction.” Chris discusses his desire to be measured, and how those measurements and that data make him feel better about himself. Self-tracking allows him to quantify his health in terms of numbers and receive satisfaction every time he improves that number in some way. As one would feel satisfied from working and seeing their bank account grow, showing the payoffs, self-tracking may be a similar mentality, as someone who works out wants to see the payoff and results from working out and putting in that effort. Amy, a female Apple Watch user states, “just getting that notification that you completed a goal is so motivating.” Mary explains, “when I go to bed and all three rings are closed I feel great and I had an active day it’s almost like a reinforcement of the good feelings.” When the rings are closed on her device, meaning she has completed each activity goal for the day, this acts as a reinforcement of the fact that she feels good about being active that day. Therefore, college students express a desire to be measured and challenged by their data flows, but also look to their data as a source of external validation that their hard work is worth it.

New users reflect on the fact that the satisfactions of self-tracking are short-lived: the data can make one feel motivated and accomplished one day, and then guilty and shameful the next. Emma, a female new Fitbit user says, “I would say I 100 percent felt more confident when I did meet all the goals but then that really crashed if I didn’t so it was a fine line.” Emma experienced the short-lived satisfaction of self-tracking, as she felt great when she reached the goals, but “crashed” when she failed to. Even if one satisfied a health-related goal one day, they have to start all over the next, resulting in an endless cycle where one never actually reached a final goal.[64] Furthermore, Emma’s experience highlights how self-tracking devices can become entangled with our sense of self. Meeting all of the goals helped her feel “confident,” affecting how she felt about herself as a person. In addition, failing to meet goals resulted in feelings of self-doubt. Therefore, these devices influence how college students perceive themselves.

Furthermore, looking at the data generated from a wearable fitness tracker can result in moving from statistical analysis to self-optimization: rather than purely seeing that data, users feel the need to make those numbers better. Amy, an experienced Apple Watch user, discusses how she receives her weekly summary from the past week at the beginning
of a new week, enticing her to “do better” than the previous week. She also explains how each weekly summary pushes the user to do better the next week: “after the weekly summary you can like change your calorie goal, so it sets it based on like your average, and so like they push you to like keep meeting your goals, so, and you can set it low if you want, but I like the way that they do it.” Amy discusses how she likes the way her Apple Watch sets her calorie goal, representing the customizable features of self-tracking technologies and how devices are tailored to “push you to keep meeting your goals.” According to Millington, “customization is important in that it is key to optimization.” This means that self-tracking device users can optimize their lifestyle of fit living as devices push users to reach certain, customized goals, tailored to how they want to transform themselves. In addition, Amy explains how her Apple Watch pushes her to improve week by week through the weekly summaries, representing how self-tracking pushes one to engage in practices of self-betterment.

Receiving feedback from a device also informs the user of their progress throughout the day, encouraging the techno-gaze so college students can reflect on their progress and improve upon it. Mary, an experienced Apple Watch user, discusses receiving feedback from her device, explaining, “It’s so weird, it like gets in your head and it’s always on your wrist, always telling you something, good or bad.” Because her watch is always “telling” her something, this encourages Mary to engage with that data and assess that information. Furthermore, the device is always telling her something “good” or “bad” which means such data can elicit different feelings based on the information she receives. Charlotte, a new Apple Watch user, explains that before using a device she would not have thought about her activity levels in the same way. She says she didn’t think about her activity “in the sense of like you didn’t do a whole lot today, so like maybe you should go to these things.” Now she cannot miss the overall picture of data and messaging involved with exercise, saying, “It’s very blatantly right there that I probably should go do something.” In this way Charlotte acknowledges the “techno-gaze” and how using a wearable fitness tracker encourages and persuades the user to engage with their data and thus act on that information. This gaze represents a form of social control because it encourages users to take in this information and thus control themselves by listening to that information.

Users of self-tracking devices also discuss how they use their devices to monitor others’ health and fitness information through technology. Amy, a female Apple Watch user discusses social connectivity with friends on her device. She elaborates on the technology’s features that allow her to check on friends’ workout statuses and compete with them. When we asked her how this makes her feel, she says:

Sometimes I will get a notification, and when I see the workout I will be like good workout, they went on a run like clearly a long run or like they’ll have like a lot of calories for whatever high intensity intervals, so I’m like that’s cool I should try that, um but sometimes like I’ll see people will have like such a good high calorie burn, or a long workout and like mine like I might not have burned a lot of calories even though it was still a good workout, so I’m like ugh shoot I should have done something else, so sometimes it like bothers me.

Amy discusses how she feels when she receives notifications about her friends’ workouts. She explains that sometimes she feels positive emotions towards her friends’ activity levels, but also finds herself comparing her own workouts to that of her friends, such as when they have a high calorie burn or a long workout. She then feels like her workout was inadequate and like she “should have done something else.” Amy’s experience represents how these devices not only push the individual to control themselves, but also heighten that sense of needing to control and monitor oneself by allowing users to share workouts and progress.

Users may feel like they need to push themselves even more when they share their activity with their friends. Mary explains her experience with friend, sharing, “My other cousin also got [an Apple Watch] and she would do Pilates at like 7am because she would work so she would be done at 7 and I wouldn’t have even woken up until like 9 and I would feel so shitty (laughs). I would feel like ahh! Damnit she’s already beating me.” The ability to monitor others
or oneself in a group context through the use of fitness tracking technologies relates to Foucault’s idea of the panopticon. The panopticon is a prison designed so that the monitoring gaze of those in power could operate centrally to observe inmates in their separate cells, which means a small number of those with authority can observe a large number of individuals. While the panopticon means that a small number of those in power can exert the “gaze” on a large number of inmates, digital health technologies complicate this idea. Devices invite the individual to turn the gaze on themselves in the context of a larger group, such as friends, family, or co-workers. Thus, Mary’s experience represents how her own health is now viewed in the context of family competition.

Agency and Opting-Out of Tracking

This section focuses on college students’ attempts at resistance and how individuals choose to “opt out” of this tracking phenomenon and/or challenge dominant discourses about the “right” way to approach health. In addition, we specifically examine agency for previous users, in order to understand how and why they stopped using a self-tracking device. Our survey data show that fitness and working out is highly valued on this elite campus, as over half of the sample (67.5%) indicate they workout often or always; however, some of our participants resist the social and environmental pressure to workout as a way to optimize the body. Mark, a male non-user, reflects on the fact that mental health is hardly prevalent in our society, saying, “I think sometimes physical health is more pressed in our society and seen as more important because it’s physical so you can see that, whereas mental health is less valued.” Mark says that he doesn’t workout at [college] in the sense of going to the gym due to a lack of time; rather, he walks to class rather than driving. Unlike many participants that felt pressure to take up fit living in college, Mark says the role of health and fitness hasn’t really changed since coming to college. He says, “It probably has stayed the same I would say, not a lot of changes, maybe a lot more aware of mental health moving from high school to college.” He describes his primary health goal as, “Get more sleep, and try to manage that because that can be a stressor.” Therefore, although there are pressures to engage in projects of self-optimization of the body in the context of an elite college environment, students are agents in the ways in which they choose to focus on health. Furthermore, Rachel, a female non-user, doesn’t conform to the local collegiate value of working out. When I asked whether she has an athletic routine, she says that she does not, and explains, “Just because I don’t enjoy it at all, and I don’t really feel like I need to workout. So if I’m not enjoying it why would I do it… working out has always been about having fun for me not the physical aspect.” Therefore, Rachel may be more focused on pleasure, and because she doesn’t associate this with exercise, she doesn’t feel the need to workout like much of the student body. Therefore, although many students place a high value on health and fitness, not all students are placing immense pressure on themselves to keep up with fitness routines.

Not all college students got hooked on self-tracking or believe these technologies promote the best version of health. Previous users provide a unique perspective where this is concerned. Carly, a female previous user explains when she received her Fitbit, “I was like oh this is great but it was more of a curiosity thing. I just never really got into it and it didn’t really do much for me, or really I didn’t do much from having it.” While some college students are drawn to quantifying their health and fitness, Carly didn’t believe that just because she was measuring her activity that her device was helping her “do better.” In addition, she decided not to use the device anymore, showing her agency in this context. She states, “Um I think I kind of recognized that I like didn’t really care that much when I hit a goal, and I wasn’t changing my behaviors at all so like I think it kind of just died one day and I’m not going to recharge it.” Carly felt ambivalent about her device and wasn’t going to blindly follow its advice and change her behaviors. Similarly, in Ruckenstein study, some participants had no visible interest in their data. Although corporations that sell these devices promote the idea that health is about getting a certain number of steps in a day or reaching activity goals, not all students really see the point in such health behaviors.

Jamie, another previous user, has a critical perspective towards self-tracking, finding it “detrimental” after witnessing her brother’s experience. She explains:

*I don’t necessarily think tracking is the best way. Like my brother does it because he lifts weights and stuff and wants to keep his muscle composition the same but I*
think it can be really detrimental. Like I said we can’t even go out to dinner with him, he’ll only get a water and it’s been difficult so I think it’s made people so like, I have to eat this, I have eat that...As someone who just kind of eats whatever and tries to eat intuitively and people that track calories, like that’s not the way your body is supposed to be, so I feel like it’s made us very, like this one pathway to getting, being fit or whatever and I think it looks different for everybody.

Jamie recognizes that tracking may not be the “best way” because it causes high levels of social control on the individual. Jamie’s brother quantifies his health, which then pushes him to control himself internally, such as only ordering water when out to dinner. Jamie shows her agency as she does not self-track and eats intuitively. In addition, she recognizes that tracking devices promote a single “pathway” towards health when in reality different people have different bodies and needs. Previous studies have found similar perspectives towards self-tracking: in Ruckenstein’s study, some participants found self-tracking as too individualistic and constraining onto everyday lives. Therefore, our findings are consistent with previous findings regarding agency in the context of wearable fitness tracking.

Furthermore, some of our new users also found that they did not like quantifying their health at the end of the three-week period. Sophie said that she felt like wearing the device was a good experience. However, she goes on to say, “I feel like [self-tracking] is a less healthy way of life, at least for me like ‘cause I’d rather have it be an individual experience and kind of personalize my own lifestyle for my own health, and not be crazed with like the numbers and stuff.” Sophie recognizes that living by numbers and quantifying one’s health may not be the healthiest way of life. Fitness devices may present one view of health and fitness that isn’t the right fit for everyone. Sophie says that after the study is over, “I can go back to not checking it all the time. I guess and it’s like if I kept wearing it I would keep tracking but like I’m glad I’m not going to feel like I have to compulsively check it.” Therefore, while some college students adopt these devices as a way to optimize the self through numbers and data, others feel as though they would rather listen to their natural selves regarding health and fitness.

Emma, the other female new user, recognized that quantifying her health through self-tracking could lead to unhealthy thoughts and behaviors. At the end of the study, her recognition of the mechanisms of social control through fitness tracking shows her ability to think independently in the broader context of the health and fitness industry. At the time of our final interview after using the device, she had already deleted the app and said, “I was happy to take it off, I thought it was really interesting to try for three weeks but it’s not something that I would want to have on me constantly... I was done with it in a lot of ways and ready to turn it back in.” Emma was unwilling to be tracked. When I asked her why she was done with it, she elaborates:

*I think now I just get more of a break now that I’m not constantly checking it and like thinking about it as much. I mean I took it off at the end of the day on Friday so it’s been off for a few days and I’ve worked out all of those days, and just working out for a certain amount of time or until my body feels done is much easier and I think better for me than doing it to a certain number of steps or minutes on the app. I don’t know it just feels more natural.*

Emma says that she is “done” with the device because she doesn’t want to be constantly checking it and thinking about her activity. While tracking devices can make our bodies more legible through measurement and observation, Emma did not want to be thinking about her body and its functions in the way that the device was prompting her too. Emma feels as though listening to her body is easier than having to work out in order to meet the goals on a device.

New users and experienced users discuss how there is less social control when they choose to take off the device. Although some participants did not express a desire to stop using the tracking device, college students still revealed agency when they took a break from actively using the technology. Chris, a new Fitbit user, discusses how he felt when he had the device on vs. off. He says, “I think the days I felt most free were the days I didn’t wear it or forgot to check it and just forgot about it.” Therefore, when Chris doesn’t have the Fitbit on, and chooses not to wear the device, he feels free from the social control that comes with using a self-tracker. Mary, a female Apple Watch user, elaborates on how if the device becomes too much, she
chooses to take it off. She says:

*I like find myself getting anxious and like sad, like I didn’t reach, I was sick and I didn’t reach my goals and I had to just take it off. I was like fuck this, I can’t have it tell me to get up one more time, I’m napping because I have a fever. It’s weird how it gets in your head and pushes you to always be moving always."

Although Mary still uses her Apple Watch to self-track, she recognizes that when the device makes her feel “anxious” and “sad,” she has the power to take it off and say “fuck this.” However, this discourse represents the high amount of social control users face when using the device, such as how they can’t even take a break from its reminders to move when sick. Although users face high levels of social control, they are agents in their decisions to ignore or take the device off when it becomes too much.

**Discussion and Conclusion**

Three weeks into using a tracking device for health and fitness, our participants felt a heightened responsibility for health. All participants discuss feelings of shame and guilt for missing the gym, calling themselves unproductive and lazy. Such internalized oppression and self-negativity reflects moral imperatives to be healthy in our society today, perpetuated by neoliberal belief systems. Those who do not conform to such standards are stigmatized. For college students who have access to a nice gym and quality food options on an elite college campus the perception is there is no excuse not to take personal responsibility for their health, while in our larger society many don’t have the resources to maintain basic health, like safe neighborhoods, quality food options, or access to healthcare.

In addition, the “techno-gaze” or seeing one’s activity levels in a datafied form creates a hyperawareness of the body’s functions and movements. Users of self-tracking devices are encouraged by the techno-gaze to engage with and confront their data. As a result, when college students see their data through this digital interface, they feel a heightened responsibility to increase their progress and ensure their goals are being met. However, without a self-tracking device, this “visualization of inferiority,” or lacking enough activity, would not be apparent. This explains why users feel that using a device heightens responsibility for health and fitness.

This research shows how self-tracking devices become entangled with our sense of self as technology re-socializes us on how to live. Fitness tracking users in our sample, especially experienced users who have been using these devices for a long time, discuss how they feel lost and blind without their device. In addition, participants discuss their feelings of anxiousness if they don’t have their device when they workout; they feel almost as if the workout doesn’t count because they can’t see it logged on the watch. These narratives relate to the theory of biopedagogy, or how these devices train and teach our bodies how to live. For example, our participants discuss “micro-practices” such as getting up and walking around the library to hit a certain number of steps per hour. Feeling lost and frustrated without a self-tracker highlights a reliance on technology and our quantification of health. As wearable fitness trackers train our bodies and minds how to pursue health, this may place individuals and society at large on a singular pathway towards well-being.

Our findings also contribute to the sociology of emotions. Fitness tracking and physical health are never separate from mental health. Self-tracking devices become inextricable from our sense of self as the data one generates can result in feelings of confidence and validation, but also guilt and shame. Participants discuss how when they hit the goals on their device, they feel more confident, and these results are a recognition of their hard work, resulting in positive feelings. Not surprisingly these devices validate those bodies and individuals that are the most fit by corporate and societal standards, and punish those that deviate from fitness norms, reinforcing existing cycles of health stratification. Ultimately, however, our research reveals how fitness trackers ultimately reinforce health insecurity for the vast majority of users.

The ways that self-tracking can influence emotions is important in the context of an elite campus environment, where students face a lot of pressure and competitiveness in both academics and peer culture, and most students are already struggling to keep their heads above water in terms of anxiety and depression. According to a study by the Pew Research Center, seven in ten teens see anxiety and depression as a major problem among their peers. In addition, 61 percent...
of teens say they feel pressure to get good grades, 29 percent feel pressure to look good, and 28 percent feel pressure to fit in socially.73 Self-tracking devices may be the icing on top of the cake in terms of further perpetuating feelings of vulnerability and self-doubt. As college students use wearable fitness trackers, it becomes a never-ending goal of satisfaction: once you achieve these goals one day, you just have to start all over the next. In addition, these devices further promote social comparison and competition among friends, especially with the Apple Watch, where you can share your workouts with others. There may be further implications regarding sharing activity and how this can heighten feelings of low confidence. Therefore, there need to be more open dialogue regarding these devices and the impact they can have on young adults in particular.

Finally, college students grapple with agency in the context of self-tracking devices. Female new users in particular discussed with us their unwillingness to quantify their health, and both were ready to return the devices at the end of the study. These students believe that following the “intuitive” self is a better way to pursue health rather than measuring workouts by a certain number of miles, active minutes, or steps per day. However, experienced self-tracker users also discuss their frustrations with the device, and how they need to take them off because they don’t want a million notifications telling them to workout when they physically cannot. Even though this is a recognition of the high level of social control, the participants still use the device.

Although college students may choose to “opt out” of the tracking phenomenon, do they ever really escape an era of self-policing? The answer is no. Even the non-user participants in this study discuss the need to workout and engage in projects of self-optimization. They all acknowledge the reality of self-surveilling in a larger culture of health self-optimization. While the non-users may not quantify and “datafy” their lives to an extreme degree, they are likely still self-policing in a variety of ways, with or without fitness trackers.

Endnotes

3. Ibid.
11. Ibid.


32. Fotopoulou and O’Riordan, “Training to Self-Care: Fitness Tracking, Biopedagogy and the Healthy Consumer,” pp. 54-68.


34. Fotopoulou and O’Riordan, “Training to Self-Care: Fitness Tracking, Biopedagogy and the Healthy Consumer,” pp. 54-68.


36. Ibid. p. 75.

37. Ibid. p. 68-84.

38. Ibid. p. 69.


40. Ibid.


42. Conner-Greene et. al, “Perceived Social Climate


46. Ibid, p. 43.

47. Ibid, p. 47.


53. Fotopoulou and O’Riordan, “Training to Self-Care: Fitness Tracking, Biopedagogy and the Healthy Consumer,” pp. 54-68.


56. Ibid.


60. Fotopoulou and O’Riordan, “Training to Self-Care: Fitness Tracking, Biopedagogy and the Healthy Consumer,” pp. 54-68.


62. Ibid.


68. Ibid.


73. Fotopoulou and O’Riordan, “Training to Self-Care: Fitness Tracking, Biopedagogy and the Healthy Consumer,” pp. 54-68.

74. Pew Research Center, “Tracking for Health.”

75. Pew Research Center, “Tracking for Health”

**Bibliography**


Fitbit, Inc. “Why Fitbit.” Fitbit.com,


Lupton, D. “M-Health and Health Promotion: The Digital Cyborg and Surveillance Society.” *Social Theory & Health* 10, no. 3 (2012): 229-244.


Smith, Gavin J. D. and Ben Vonthethoff. “Health by Numbers? Exploring the Practice and Experi-


The Buck-Passer’s Dilemma: Evidence from Japan

Matthew Salah, Swarthmore College

Abstract

Scholars have long noted states’ tendency “to pass the buck” when asked to contribute to international public goods. The security arena is no exception. However, many have extended buck-passing theory to account for some states’ decisions to neglect their own military development in favor of free-riding on a powerful patron state. This paper questions this explanation, showing that security-maximizing states should be skeptical of completely outsourcing their own protection. The paper then turns to Japan, a country many contend is free-riding on its U.S. alliance, to demonstrate how the buck-passing theory fails to explain Japan’s restrained military record. The result is that the buck-passing explanation fails on both theoretical and empirical grounds: first, highly dependent states should find it difficult to overcome the fear of allied abandonment, and second, when states do manage to retain weak or otherwise constrained militaries, it is largely due to normative rather than strategic considerations. This has pivotal implications for U.S. foreign policy and international relations scholarship. It suggests that the U.S. allies that have remained militarily weak have not simply chosen to outsource their security to the U.S., but rather have avoided military development for other normative or institutional reasons. And it suggests that scholars should look to paradigms other than realism to explain the behavior of these and other security outliers.

Introduction

The election of Donald Trump in the United States in 2016 sent shockwaves through the U.S.-Japan alliance, which was formally established in 1951 with the signing of the U.S.-Japan Security Treaty. Trump has denounced the alliance as unfair, has reportedly privately considered ending it, and on the campaign trail, even suggested that Japan should acquire nuclear weapons.¹ This has reignited long-standing fears that Japan will become increasingly militarily assertive in East Asia to the detriment of regional stability. In the aftermath of Trump’s election, debates within Japan about the role of the military have re-emerged, with Prime Minister Shinzo Abe recently pushing to revise Article 9 of Japan’s anti-military constitution and insisting that it does not explicitly prevent nuclearization.² Yet, so far, Tokyo has remained surprisingly hesitant to fortify its military or further project power. In fact, the Japanese government immediately reaffirmed its commitment to non-proliferation following Trump’s remarks—an affirmation consistent with Japan’s largely pacifist security record since occupying U.S. forces installed its “Peace Constitution” after World War II. Today, while Japan is undoubtedly a serious player on the world stage, its military role in global affairs notably pales in comparison to its economic might due to substantial restrictions imposed by the government on the nation’s military policy.³

The realist paradigm characterizes the international order as continuously prone to conflict, inherently dangerous, and based largely on self-help.
How can militarily weak states such as Japan persist in such a system, especially when they constrain their militaries by choice? Realists claim to have an answer: buck-passing. When asked to provide public goods for the benefit of the entire international community, states will often “pass the buck”—shift responsibility—to others, hoping to free-ride on their contributions. This occurs not only in arenas such as environmental protection or pandemic preparedness, but also with regard to security. Rather than building up their own military, states may count on their allies to bear the burden of maintaining the balance of power. In the most extreme case, as many argue applies to Japan, one state may consistently underinvest in its own military, entrusting its defense to a more powerful ally. This form of buck-passing features a client state—the buck-passer—and a patron state to which the client outsources its security.

While the patron-client buck-passing story has intuitive appeal—states would certainly appreciate free protection—it presents an immediate predicament. I will call this predicament “buck-passer’s dilemma”: how can the buck-passing state trust its patron to come to its defense in the event of a foreign attack? After all, for the patron state, responding militarily may be an unacceptably costly option. Charles de Gaulle feared this exactly dynamic when he asked if the U.S. would truly sacrifice New York to save Paris from a Soviet march through Europe. The core “buck-passer’s dilemma,” then, is whether the client state can truly risk outsourcing its security knowing that its patron may lose interest in protecting it. As John Mearsheimer writes, “buck-passing is not a foolproof strategy... Its chief drawback is that the buck-catcher [the patron] might fail to check the aggressor, leaving the buck-passer in a precarious strategic position.” A realist approach to international relations must be capable of explaining why states would be willing to take this gamble in a self-help world. If states must prepare for the worst intentions of others, as realism centrally posits, then buck-passing appears foolish at best.

This paper argues that the patron-client buck-passing theory fails to explain state behavior through two contributions to the existing literature. First, it analyzes the strategic logic of buck-passing—specifically the patron-client buck-passing relationship—and draws on balance of power theory to introduce the “buck-passer’s dilemma.” In theory, realism implies that states should hesitate to build any meaningful dependency on others. The buck-passer’s dilemma, therefore, should be very hard to overcome. Second, I offer an in-depth inspection of one of the (supposed) paradigm cases of buck-passing: Japan in the post-World War II era. At first glance, Japan’s behavior exhibits clear signs of buck-passing: it has shored up a robust security arrangement with the U.S. while keeping its military forces small (over the last decade, Japan dedicated a mere 2.5% of its government budget to its military, while many of its neighbors commit over four times that percentage) and incredibly constrained (Article 9 of its constitution, for example, prohibits the use of offensive force). Thus, Japan represents a hard case for the argument that buck-passing theory cannot explain states’ behavior. Yet, as I will argue, the buck-passing explanation seems largely inconsistent with the specifics of Japan’s foreign policy record and overlooks the role of domestic anti-military attitudes in explaining Japan’s military restraint. Indeed, many scholars contest the buck-passing account Japan’s military policy, citing everything from domestic norms to institutional inertia to Japan’s “internationalizing economic model” as alternative explanations for Japan’s restraint.

This paper contributes to the ongoing debates over Japanese foreign policy by showing that the buck-passing framework’s predictions for a more assertive and militaristic Japan following the Cold War come up short. Most importantly, the paper couches those debates in the broader theoretical disputes around balancing behavior, arguing that the buck-passing theory’s inapplicability to Japan indicates a broader flaw in its logic: the buck-passer’s dilemma renders its story internally inconsistent. In short, the buck-passer’s dilemma should be difficult to overcome, and while Japan has managed to overcome it, it is due to its anti-military security identity (a normative factor) rather than its position in the international distribution of power (a structural variable).

The first section, buck-passing and balance of power theory, outlines buck-passing’s logic from a theoretical standpoint. It begins with a literature review, then summarizes the strategic logic behind buck-passing, and finally introduces the buck-passer’s dilemma to underline the failings of the buck-passing narrative. The second section focuses on the case study of Japan. It outlines how buck-passing theory has been applied...
to Japan, and then demonstrates the salience of the buck-passer’s dilemma in Japanese foreign policy. It then advances the normative theory of anti-militarism as an alternative explanation for Japanese restraint. Finally, it tests the buck-passing theory against Japan’s empirical record in the 1990s: an era where the predictions of the buck-passing and anti-militarism explanations sharply diverged.

I conclude that the realist theory of buck-passing fails to explain postwar Japanese military restraint. Instead, a normative approach focusing on domestic anti-militarism offers a much more rich and accurate account of Japanese security policy. There is little reason to expect that buck-passing theory’s inadequacy is isolated to the case of Japan. Rather, its inapplicability to Japan indicates a broader flaw in its logic: the buck-passer’s dilemma should be incredibly difficult for any state to overcome on strategic grounds.

**The Theory of Buck-Passing**

*Balance of Power Theory: Buck-Passing in Context*

Many of the seminal works in international relations have analyzed states’ tendency to rely on alliances to maintain security. External balancing—the formation of coalitions to counterbalance against a rising power—is such a well-documented phenomenon in modern European history that it is often treated as an “iron law” of international affairs. External balancing must be distinguished from internal balancing, which refers to states’ attempts to build up their own domestic military power. Moreover, as Stephen Walt famously observed, balancing—both internal and external—must also be distinguished from bandwagoning, where states form an alliance with rather than against the most powerful state in the system.

In this context, buck-passing refers to shirking on one’s own commitments to counterbalance rising powers and instead relying on others to carry the burden of military development. Maintaining the balance of power, after all, is a public good: all states benefit equally from a stable international order, no matter how many resources they invest in securing it. The practice of shifting collective military burdens to others has been dubbed “underbalancing” by Randall Schweller. Within the North Atlantic Treaty Organization (NATO), for instance, quantitative evidence from Thomas Plumper and Eric Neumayer suggests that most small members free-ride on more powerful states’ contributions.

Importantly, states can buck-pass whether they have chosen to balance or bandwagon: they can rely on others to prevent a rising power from achieving hegemony (in the balancing case), or they can hope to free-ride on its ally’s enormous financial commitments towards securing regional hegemony (in the bandwagoning case). Japan is a particularly interesting case since its buck-passing behavior shows signs of both balancing and bandwagoning; it has clearly bandwagoned on U.S. global hegemony, but in its own region, it has historically relied on the U.S. to balance against rising powers, especially the China and the Soviet Union. The factor distinguishing buck-passing, therefore, is not whether the state balances or bandwagons, but whether they exhibit a clear preference for external alliances over internal military development.

While balance of power literature often neglects to make explicit distinctions between external and internal balancing, when such distinctions are drawn, realists tend to conclude that states prefer internal balancing. Given the scarcity of trust in realists’ depictions of the international system, Eric Heginbotham and Richard Samuels observe that structural realism broadly accepts “that states prefer to maintain independent military capabilities.” Kenneth Waltz has noted that states tend to find internal balancing more reliable due to endemic mistrust of their allies. Similarly, in an extensive review of the last 200 years of alliance formation, Joseph Parent and Sebastian Rosato conclude that “great powers have generally doubted the reliability of their allies and of their opponents’ allies.”

Those that believe states prefer external balancing tend to rely on variables that diverge from neorealism’s overwhelming emphasis on international structure: Randall Schweller, for instance, has noted that internal balancing has higher domestic political costs because it requires mobilizing more resources.

Despite most states’ apparent preference for internal balancing, external coalitions invariably emerge to different degrees. As Dominic Tierney observes, alliance systems have often been distinguished by their “tightness”: tight alliances are synonymous with chain-ganging—one alliance member dragging others towards aggression stance or war—while loose alliances represent buck-passing behavior—each state is hes-
tant to commit fully to the alliance, hoping to make others bear the military burden. Thomas Christensen and Jack Snyder analyze the conditions under which alliance systems feature chain-ganging or buck-passing, and argue that a defensive military advantage prompts buck-passing behavior by lowering the costs of underbalancing.

In our current unipolar world, debates surrounding buck-passing have become all the more important: since the U.S. alliance system extends across the globe, Washington needs to know how states will react to changes in its military policy. Kurt Campbell argues that U.S. military activity represents the single “most important ingredient” in determining whether U.S. allies, including “Japan, South Korea, Taiwan, Germany, Egypt, and others” will pursue militarism and nuclear proliferation. Similarly, Rory Medcalf worries that a decline in U.S. military presence in Asia will force U.S. allies in the region to take their security into their own hands and even acquire nuclear weapons; Stephen Ellis argues that U.S. power in East Asia is “a critical factor in explaining why Japan, South Korea and Taiwan have not become nuclear actors.” In the case of Japan specifically, Satake Tomohiko argues that Japan’s commitment to nuclear non-proliferation “depends largely, if not solely” on U.S. military commitments and capabilities. These are bold claims that need backing up. If the buck-passing theory holds, then many of these predictions will too: the absence of a firm U.S. commitment will force the states currently free-riding on U.S. power to take matters into their own hands, potentially triggering a global nuclear cascade. But if the buck-passing theory is flawed, as I will contend, then there are other factors inhibiting such states from militarizing, and one should read these alarmist warnings as yet another way of justifying bloated U.S. military spending.

The Strategic Logic of Buck-Passing

The central premise of the patron-client buck-passing strategy is that a state would rather free-ride on the security commitments of its allies than bear the costs of militarizing itself. A state achieves two key benefits from shifting responsibility for its security. First, it reduces the economic burden of maintaining a sizable military. These savings may even improve the states’ long-run security outlook by giving the government more fiscal flexibility to pursue various military enhancement strategies in the future. Second, it helps temper the security dilemma. Since the buck-passer itself maintains little offensive capability, other states will be less inclined to balance against it. Moreover, alliance commitments are far more credibly called upon in defensive situations: a state may feel compelled to support an ally that is defending itself from attack but not one that has lashed out aggressively. Indeed, many formal alliances, such as the North Atlantic Treaty Organization (NATO), do not obligate security assistance when a state in the alliance initiates conflict. Thus, a buck-passing strategy can readily be distinguished as defensive, allowing the buck-passer to maintain sufficient security without threatening others.

However, buck-passing is just one of many potential strategies, so a complete theory of international politics must identify the conditions that lead a state to buck-pass. Two structural factors have been identified as conducive to buck-passing behavior. First, the buck-passer must have a powerful patron committed to its defense. Absent an ally sufficiently capable of and interested in defending the buck-passing state, neglecting its own military power would it far too vulnerable. Second, a defensive military advantage lowers the risk of buck-passing because it increases the costs that aggressors face and improves the chance that minimal efforts will stave off an invasion.

Ultimately, buck-passing is a gamble, and the relative strength of these two variables should at least partially determine a state’s commitment to a buck-passing strategy. When a state faces few foreign threats, buck-passing becomes more appealing: when its threat environment deteriorates, it should pursue rearmament. Likewise, when it has faith in its patron’s promises of military support, a buck-passer should feel comfortable relying on the patron’s protection, while doubts in the alliance should pressure it to increase its own military capabilities. Given the scarcity of unadulterated trust in a realist world, this variable is perhaps better coded as the extent to which security threats to the buck-passer also threaten its patron’s interests: the client will have higher “trust” when the patron has strong incentives to back it up.

Once a state has chosen to buck-pass, its strategy entails a number of interdependent agendas. First, a buck-passer will take steps to secure its patron’s commitment. Abandonment is the biggest
risk a buck-passing state faces; without support from a powerful patron, there is nothing left to prevent other states from exploiting its military weakness. A buck-passer can shore up its alliance by contributing to its patron's economic, military, or diplomatic initiatives, or by threatening undesirable behavior should the patron defect.

Second, a buck-passer will try to maintain a minimum level of deterrence, such as having a capable ballistic missile program or maintaining nuclear breakout potential, without expending too much cost or appearing threatening. The combination of a secure alliance and limited autonomous deterrence helps a buck-passer achieve what's known as “threshold deterrence”: the buck-passer retains sufficient capabilities to deter actions too insignificant to prompt its patron to intervene, and it can rely on the strength of its patron to deter larger incursions.31

Finally, a buck-passer will make the minimum possible contribution to its defense without jeopardizing the first two endeavors.32 Spending too little risks eroding threshold deterrence and may introduce frictions into the alliance if the state is perceived as free-riding, but spending more than the minimum necessary amount undermines the original purpose of buck-passing.

The Buck-Passer’s Dilemma

The central dilemma any buck-passing state faces is that of trust: how can the buck-passer have enough faith in its patron’s commitments to outsource its security? Because the stakes of an ally defecting are potentially existential, states must plan for the worst-case scenario and ensure their survival independently of their alliances. Although a buck-passing state may have enough confidence in its alliance(s) to mitigate the fear of outright defection, the buck-passer can never be completely certain that the patron state will intervene once conflict breaks out: intervening may often conflict with the patron’s immediate interests. After all, the buck-passer always maintains a much stronger interest in repelling an attack on its own soil than the patron does, so a rationalist framework would expect the patron to under-provide security compared to buck-passer’s preferred amount. In a self-help world, alliance commitments are bound to falter when the security environment changes, so realists—including Waltz, Heginbotham, Samuels, Parent, Rosato and others—overwhelmingly contend that states tend to find internal, rather than external balancing more reliable. The potential for domestic political changes within the patron state should also undermine a buck-passer’s confidence in its security: a sudden change in popular attitudes or leadership could disrupt the alliance before the buck-passer has time to properly rearm, leaving it incredibly vulnerable.

Neither of the supposed advantages to buck-passing seem compelling enough to overcome this inherent risk. The claim that buck-passing saves money seems plausible, but few realists believe that states, under any circumstances, will sacrifice their security for marginal economic gains. Buck-passing also has questionable long-term utility as a reassurance strategy, as states know the buck-passer can choose to remilitarize at any moment. Further, relying on an outside military power is unlikely to provide enough security for a state to forsake other, more conventional forms of balancing. A robust security pact is certainly one tool a state can use to enhance their defensive capabilities, but given states’ endemic mistrust of others’ commitments—including their allies’—an alliance alone seems insufficient. An alliance may contribute to a sense of security, but it cannot explain why some states have chosen to outsource their military power to such an extreme extent.

Nor can the presence of a defensive military advantage adequately explain why states have taken the dangerous gamble of buck-passing. When Thomas Christensen and Jack Snyder wrote that defensive advantages encourage buck-passing, they referred to states passing the buck when it came time to act on their alliance commitments—not states who chose to outsource their own security to an ally.33 In the case of a client-patron relationship, Christensen and Snyder’s logic imagines the patron as the buck-passer, dragging its feet on defending its client because any aggressor would either be repelled by the client alone or weakened by the outsized costs of the invasion to the point that they no longer pose a serious threat to the patron. Here, the patron’s defensive advantage actually serves to impair the client’s confidence in the alliance by reducing the patron’s interest in coming to the client’s defense.34

Thus, while the buck-passing explanation has some intuitive appeal, its inherent risks should be dif-
ficult for states to overcome. In the next section, I turn to a case study of Japan to demonstrate the empirical salience of the buck-passer’s dilemma and show that Japan’s ability to overcome it stems from normative rather than strategic concerns.

**Case Study of Japan**

**A Security Outlier**

Japan’s history as a militarily restrained nation begins in the aftermath of World War II. During the U.S. occupation of imperial Japan, the allied powers imposed a variety of anti-war constraints on the newly-democratic country, including a prohibition on weapons production and a clause in Japan’s constitution, known as Article 9, which outlawed war as means for settling international disputes. While Japan had no choice but to accept these impositions during the U.S. occupation, once the occupation ended in 1952, Japan retained many of these regulations and even voluntarily added further constraints to its military policy. In 1954, Japan did establish a de facto military, the Self-Defense Forces (SDF). However, unlike typical armed forces, the SDF did not constitute a full-fledged army and its original mandate only permitted the use of force to defend from a direct attack on Japan. Soon after, the Diet (Japan’s bicameral legislature) passed a ban on the overseas dispatch of the SDF and prohibited the acquisition of certain offensive weapons, such as long-range bombers. Japan enacted a host of further restrictions in the following decades. In 1967, Prime Minister Eisaku Sato formalized a ban on arms exports, prohibited the military use of space, and articulated Japan’s Three Non-Nuclear Principles of non-possession, non-production, and non-introduction. In 1976, Prime Minister Miki Takeo informally committed to a one percent of GNP cap on military spending and added additional provisions to the arms export ban. In recent years, however, some have concluded that Japanese security policy has begun to normalize. This position has some intuitive appeal, since, in terms of the raw force size and quality, Japan ranks among the world’s top military powers. It has as many modern fighter jets, early warning systems, and well-trained pilots as the world’s most advanced air forces, and has stronger sea control capabilities than many great powers.

But it would be wrong to assume that Japan’s immense power means that it is no longer a security outlier. First, Japan’s military remains weak in many key areas. It is fully incapable of staging offensive operations, for example, limiting Japan’s ability to deter foreign attacks. These deficiencies are especially surprising relative to Japan’s economic power. Over the last decade, Japanese military spending has averaged just 2.5% of total government spending while nearby countries spend much more: South Korea spends 12.7% of its budget on the military, Taiwan, 10.3%, China, 6.7%, and Russia its 11.5%. Out of the 151 countries in SIPRI’s military expenditure database with data for 2018, Japanese spending as a share of government budget ranked 125th. As even those who see Japan as a normal military power will admit, measuring military spending relative to economic size rather than in total best gauges a state’s commitment to its military.

Second, Japan retains far too many restrictions on its security policy to consider it a normal military power. Japan still largely refrains from exporting weapons, and it retains severe restrictions on when the SDF can be deployed and the type of weaponry it can possess. Many of the recently loosened restrictions represent incremental reform rather than a complete abandonment of military restraint. For instance, in the 1980s, while the hawkish Prime Minister Yasuhiro Nakasone was able to gain approval to send the SDF on fact-finding or monitoring missions, his efforts to dispatch minesweepers to the Persian Gulf were blocked. Later, when Japan joined the U.S. war on terror, it passed laws that prevented the SDF from engaging in active combat or supplying weapons, and required Diet approval for each mission. Clearly, Japan is no ordinary military power: while many of Japan’s anti-military policies have eroded over time, Japan’s military faces some of the most substantial political constraints in the world and remains unusually weak given Japan’s economic power.

**Has Japan Passed the Buck?**

Realists have unsurprisingly attributed Japan’s behavior to buck-passing, although considerable variation remains within realist explanations of buck-passing theory. Many scholars’ explanations rely on modified versions of neorealism: Eric Heginbotham and Richard Samuels draw on mercantile realism, which
accepts that states may trade-off military interests for techno-economic ones, Tsuyoshi Kawasaki employs a postclassical realist framework, which sees states as sensitive to the economic costs of defense, and Paul Midford builds on Walt's balance of threat theory. Overall, however, buck-passing is the materialist explanation that most closely matches Japan's security record over the past 70 years, and almost all realist approaches tend to converge on some variation of the buck-passing hypothesis.

At first glance, the buck-passing explanation for Japan's military restraint seems compelling: the variables that are typically theorized as conducive to a buck-passing strategy all apply to Japan. First, on the economic front, Japan's alliance with the U.S. during the Cold War allowed them to forgo the financial strain of matching Soviet military modernization and expansion efforts. Even after the demise of the Soviet Union, U.S. protection afforded the Japanese government considerable savings. For example, in 1993, it would have cost Japan an estimated 200 billion dollars a year for a decade to build a military commensurate with its economic power.

Second, Japan's imperialist history leaves it in desperate need of a reassurance strategy. Because of Japan's record of invading and colonizing other Asian countries, regional powers have an inflated sensitivity to Japanese assertiveness; any sign of Japanese rearmament could prompt other states to balance it, sparking a regional arms race. Relying on U.S. extended deterrence allows Japan to remain secure while focusing on defensive capabilities, alleviating the security dilemma in East Asia.

Third, Japan has a powerful ally that is committed to its defense. Throughout the Cold War, the U.S. had overwhelming incentives to protect Japan, because without the alliance, U.S. policymakers feared a Soviet first strike or invasion of Japan. The politics of the Cold War, therefore, enabled Japanese buck-passing by reassuring Tokyo that the U.S. would not defect. Fourth, Japan's defensive advantage lowers the risks of buck-passing. As an island nation, Japan has a natural defensive advantage: scholars of offense-defense theory have long observed that ocean borders are more secure than land borders. Moreover, modern military technologies, such as ground-based tactical air assets, which enable control over local airspace, and antishipping cruise missiles, which aid maritime denial strategies, make crossing oceans more difficult, enhancing the defensive advantage of island states.

Finally, Japan has also pursued many of the strategies that scholars would expect from a buck-passer. Japan discourages U.S. abandonment by contributing financially and diplomatically to U.S. endeavors and reinforces the alliance symbolically by staging joint military drills. Many have also argued that Japan's extensive nuclear energy program fits this goal: by maintaining a nascent nuclear capability, Japan decreases the risk that the U.S. will defect on its alliance commitments because the U.S. fears the destabilizing effects of a nuclear Japan. Japn's nuclear breakout potential and ballistic missile capabilities also help it to maintain a minimum level of deterrence absent U.S. commitments.

Facing the Buck-Passer's Dilemma

Despite the evidence in favor of the buck-passing explanation, buck-passing actually represents a rather poor strategy for Japan. Historically, Japan has not appeared to have enough trust in U.S. commitments to justify outsourcing its security. After the outbreak of the Korean and Vietnam wars, Japan concluded that the U.S. could not deter conventional conflict, and even if the U.S. retaliated after a nuclear attack on Japan, the damage would be done. A 1994 poll showed that 85 percent of Japanese people thought at least one country posed a military threat to Japan, but polling a year later showed that only 49 percent of Japanese people thought the U.S. would assist Japan if it faced attack. Even if public opinion is an inexact measure of Tokyo's trust in Washington, this does not reflect enough confidence for Japan to pursue a risky free-riding strategy.

A defender of the buck-passing theory might respond that Japan's nuclear breakout potential minimizes the probability of U.S. abandonment to near zero, and also ensures that Japan could quickly attain sufficient deterrence absent the U.S. But even with its nascent nuclear capability, Japan will still have uncertainty over the United States' level of commitment. Because the buck-passing strategy implies that Japan prefers the status-quo over nuclearization, the U.S. may doubt Japan's willingness to go nuclear given the potential costs in terms of economic sanctions and diplomatic isolation. Even after U.S. abandonment,
Tokyo may conclude that it prefers to win back American support by playing friendly and showing that it can be a reliable partner. Furthermore, a more realistic scenario for U.S. abandonment is not a sudden withdrawal from the security pact, but instead a failure to back up Japan in a serious crisis. In such an event, the U.S. may continue to tell Japan that it will intervene on its behalf, but Japan cannot guarantee the credibility of such promises. The uncertain logic of these situations should give Japan at least some doubts about U.S. credibility and the prospects for their security absent U.S. support. After all, realism encourages states to hedge against worst case scenarios.

The second argument, that Japanese nuclearization can quickly achieve sufficient deterrence, is similarly unconvincing. Japanese nuclear breakout would be rapid but not instant: an abandoned Japan would still face several months of heightened insecurity. Once Japan faces a pressing military threat, it would be too late for Japan to develop a nuclear capability to deter its adversaries. Furthermore, nuclear weapons may deter an attack on the Japanese mainland, but they have questionable utility in defending broader interests. Could Japan credibly threaten to push the nuclear button following a Chinese incursion on its claimed territory in the East China Sea? Breakout potential does very little to secure these and other core interests. Neither of the supposed advantages to buck-passing seem to apply to Japan—at least not to the extent necessary to justify the buck-passing gamble. Buck-passing may have saved Japan a moderate amount of money, but Japan’s actions do not appear to be motivated by the prospect of economic gain. Japan’s restrictions on overseas arms sales, for instance, cannot be explained by economic interest, because the restrictions prevent them from capitalizing on a lucrative market in which the technological powerhouse could become a serious player. Furthermore, Japan’s reliance on the U.S. is more extreme than most countries would accept, which has weakened its bargaining position in bilateral economic disputes.

The logic of buck-passing as a reassurance strategy is similarly flawed. First, Japan’s low military spending historically did very little to quell its neighbors’ fears of a Japanese resurgence, since others knew that the constitution could be amended and military spending increased if Tokyo decided to change its approach. Second, it seems unlikely that China and Russia, Japan’s most powerful neighbors, will be less alarmed by a U.S.-backed Japan. If anything, Japan’s alliance with the world’s only military superpower, which has often taken a hawkish approach towards China and Russia, will make Japan’s relations with the two regional powers even more adversarial. Undoubtedly, China and Russia would feel incredibly threatened by a nuclear-armed Japan, but their regional interests would be less frustrated by a conventionally armed Japan that pushes away from the U.S., even if it means a revamped Japanese military. Third, if Japan can overcome the security dilemma by freeriding on U.S. commitments, why haven’t other U.S. allies pursued the same strategy? The United Kingdom has a similar geographic position as Japan, less nearby threats, and an even more robust U.S. alliance, but has militarized far more than Japan over the last 70 years. Since most states should want to ensure their own security without provoking balancing, it seems peculiar that Japan stands out as one of the few states to buck-pass in this way.

Japan’s defensive advantage also cannot adequately explain its particular form of buck-passing. Many have pointed out that Japan’s natural geographic security is easily overshadowed by the number of nearby threats it faces, including Russia, China, and North Korea. Southeast Asian countries and European countries face comparatively fewer direct threats, but have less of a tendency to free-ride. Moreover, Japan’s position on the frontlines of any potential attacks should make them an unlikely candidate for free-riding, since it would bear the highest costs of a potential conflict. As much as the U.S. wanted to prevent a Soviet invasion of Japan, for example, the threat of such an attack was existential for Japan and merely political for the U.S.

Anti-Militarism: An Alternative Explanation

Since restraining its military does not clearly serve Japan’s strategic interests, non-security concerns provide a better account of Tokyo’s motivations. Indeed, Japan’s security decisions seem far more consistent with the constructivist approach of anti-militarism, which emphasizes the role of World War II in shaping Japan’s cultural aversion to military power. In postwar Japan, the exogenous shock of U.S. occupation and the new political leaders it empowered forced
Japan to reconceptualize its identity. Japan’s reformed narrative blamed the military for the traumatic results of World War II. Old elites pushed this view to avoid backlash to their nationalist politics, and since it was unthinkable to blame the emperor, only the failures of military institutions could explain the atrocities. Shifting responsibility to the military was convenient for Japanese civilians as well, since it did not make them feel complicit in the war effort.

These perceptions facilitated the adoption of the so-called Yoshida Doctrine, named after then Prime Minister Shigeru Yoshida, who advocated for close security ties with the U.S. to allow Japan to focus its energy on economic pursuits. As Japanese prosperity increased, Yoshida’s position gained popularity, and people became less willing to tamper with the institutions that enabled it. Over time, the Yoshida Doctrine has become deeply embedded in Japanese identity; Japanese people see their national character as uniquely enabling a peaceful and prosperous nation.

Although Japan’s security identity does not have complete consensus support and does not fix outcomes, it nonetheless plays a powerful role in shaping Japan’s foreign policy. Japan’s security identity does not reject war flat out, but rather focuses on constraining the Japanese military; thus, it is better described as anti-militarism than complete pacifism. It consists of three facets: first, the illegitimacy of using force, except in self-defense, second, an aversion to Japanese participation in foreign wars, and third, a prohibition on the development of traditional armed forces in Japan.

Anti-militarist norms affect Japan’s military through two avenues: public opinion and institutionalization. Japanese public opinion routinely prefers passive stances to active ones and favors limiting military spending. Nearly half of Japanese people consider war illegitimate even in response to an attack; over 85 percent think that war can be avoided through cooperation. These anti-war attitudes stand out compared to other countries, and are far more optimistic than realists would recommend. Importantly, these opinions have remained consistent across generations. Public resistance to military action ultimately influences policy by preventing militarist leaders from winning election and by imposing political costs on leaders who hope to strengthen the SDF. For instance, in the late 1950s, the government’s attempts at a military revival were met with protests that ultimately removed them from power. More recently, Prime Minister Abe has announced his intention to revise Article 9 by the end of 2020 (merely to legitimize existing SDF operations), but public and intra-party resistance make this campaign an uphill battle.

In addition to the effect of public opinion, Japanese security policy is constrained by the institutionalization of the anti-militarist security identity. First, a series of formal rules prevent rearmament; for instance, the aforementioned Article 9 prohibits the development of traditional armed forces and requires two-thirds parliamentary support to repeal. Second, anti-militarism has been institutionalized through informal boundaries that inform public deliberation. Most notably, the one percent of GNP cap on military spending set by Prime Minister Takeo in 1976 has become a public symbol that defines the contours of acceptable military policy. Even though the one percent limit is non-binding, when Prime Minister Nakasone exceeded it slightly in 1985, he was met with intense backlash. Since then, Japan’s military spending has never exceeded one percent of GDP. Third, the administrative structure of Japan’s government inhibits rearmament efforts. Japan has robust civilian control over the SDF and the Japan Defense Agency (now the Ministry of Defense) has very limited institutional autonomy, with civilian institutions such as the Ministry of International Trade and Industry and the Ministry of Foreign Affairs overseeing its operations.

Why Buck-Passing Cannot Explain Japan’s Behavior

The buck-passing theory fails to explain Japan’s behavior because it overlooks the centrality of these domestic political attitudes and institutions in shaping Tokyo’s foreign policy. Indeed, the notion that Japanese militarism was tempered by the U.S. alliance seems pe-
culiar given the state of Japanese politics. Throughout postwar Japan, it was the pacifists who rejected both militarism and the U.S. alliance, while the center-right favored expanding military capabilities and integrating Japan into a more formalized Pacific Pact modeled after NATO. By 1960, a majority favored neutrality over aligning with the U.S. by nearly a two to one ratio, and when the conservative government agreed to a new security treaty, activists reacted by staging the largest public protests in Japanese history. When Prime Minister Nakasone, one of Japan’s more hawkish leaders, decided to bolster U.S.-Japan security cooperation, he accompanied it with an increase, rather than a decrease, in military spending.

Far from feeling reassured by the U.S. alliance, the Japanese public overwhelmingly considers it frivolous: a 2015 NHK poll revealed that a mere 10.3 percent of respondents considered the U.S. nuclear umbrella necessary, while 49 percent deemed it unnecessary. In fact, more Japanese people credit the “peace constitution” and the trauma of World War II for Japan’s peaceful record than U.S. security. Realists may respond that this simply proves that the pressure to maintain security overwhelmed public opinion—a strategy that was both pacifist and neutral may never have succeeded. However, it does cast doubt on the buck-passing theory that the parties and politicians most in favor of the U.S. alliance were also most likely to favor military spending increases, while the parties pushing military cutbacks and constraints also rejected deepening alliance commitments. If Japan is simply seeking a free ride, its public and leaders do not always seem aware of it.

Even if one grants that Japanese foreign policy roughly follows a buck-passing strategy, anti-militarist attitudes, and not strategic concerns, may best explain Japan’s decision to buck-pass. The legacy of World War II, the atomic bombings, and U.S. occupation has taught Japan that aggression does not pay, socializing the country towards defensive strategies. Indeed, Japanese public opinion often opposes foreign involvement on the grounds that it will increase the likelihood of war. While this could represent a rational objection to the potential costs of militarism, the prevalence of these views in Japan compared to other countries suggests that such objections represent a cultural phenomenon stemming from anti-military sentiments. Defenders of the buck-passing theory reject this argument on the grounds that immediately following World War II, Japan had an even more pacifist strategy available: the postwar left advocated that Japan remain neutral and push away from the U.S. But Japan need not have selected the most pacifist option for their strategy to be informed by anti-military culture. In fact, Japanese anti-militarism may have made buck-passing a more viable and effective strategy. Japanese leaders often used normative and institutional constraints on their military to resist U.S. calls for rearmament and refute perceptions of freeriding. For example, when diplomats from the Lyndon Johnson administration pressed Prime Minister Sato a larger Japanese presence in Southeast Asia, they understood that Japan’s political climate ruled out military options, and instead asked Sato for a greater economic contribution to the region. Other internal documents from the U.S. Bureau of the Budget further suggest that the standard buck-passing story has the causality flipped: the U.S. was historically reluctant to reduce its military contributions to Japan because anti-militarist attitudes would prevent Japan from picking up the slack, resulting in an overall deterioration of anti-communism’s position in Asia. Thus, even if one concludes that Japanese behavior roughly approximates a buck-passing strategy, domestic attitudes may nonetheless represent the underlying causal force.

Yet these objections could be discounted as theoretical nitpicking if the buck-passing explanation still roughly matches Japan’s security record. In the next section, then, I test the realist predictions for Japan against its empirical record in one of the most transformative periods in its modern military history: the early 1990s.

Can Buck-Passing Explain Japan’s Behavior in the Early 1990s?

One of the most difficult tasks in evaluating a theory of Japanese foreign policy is testing its explanatory power against the empirical record. In this section, I compare Japanese security policy in the immediate aftermath of the Cold War to predictions generated by the buck-passing theory. I chose this period for several reasons. First, the 1990s saw Japan’s first formal abandonment of an anti-militarist constraint on the SDF: the prohibition on overseas dispatch. If the buck-passing theory has any utility, it should be able to explain
this policy reversal. This period should represent a hard case for the constructivist explanation because realists have criticized its inability to explain pro-military shifts in Japanese policy. Second, the post-Cold War era gave rise to some of the clearest disagreements between the predictions generated by the buck-passing and anti-militarist theories. The early 1990s marked a potential turning point, with many realists envisioning a more heavily armed Japan. Constructivists, on the other hand, largely foresaw a continuation of the status-quo. Finally, perhaps as a result of the second factor, a wealth of literature analyzes Japanese foreign policy during this period.

To test buck-passing and anti-militarism in the post-Cold War era, I first outline the expectations of the buck-passing framework, and then compare that prediction to the empirical record. I generate the expectations of the buck-passing theory through two interrelated methods. First, I analyze realist literature immediately following end of the Cold War to uncover what contemporary scholars predicted from Japan. Second, I undertake my own cost-benefit analysis of Japanese foreign policy decisions assuming that Japan followed a buck-passing strategy. Both methods largely agree that Japan should have pursued serious rearmament in this era. I then examine Japan’s empirical foreign policy record through a process-tracing method. Most of the evidence for this section comes from three books on Japanese military history: Andrew Oros’ Normalizing Japan: Politics, Identity, and the Evolution of Security Practice, Glenn Hook’s Militarisation and Demilitarisation in Contemporary Japan, and Patrick Boyd and Richard Samuel’s Nine Lives?: The Politics of Constitutional Reform in Japan. Drawing on these sources’ account of Japanese politics in the 1990s, I trace the process by which Japanese foreign policy transformed in order to test whether it meets the expectations of the buck-passing theory.

The Expectations of Buck-Passing Theory in the 1990s

Based on the changes to Japan’s security environment in the aftermath of the Cold War, both offense and defensive realists overwhelmingly expected Japanese rearmament. As early as 1983, Mike Mochizuki predicted that changes to Japan’s security environment would undermine faith in U.S. power and resolve, leading to a surge of military realists calling for increased Japanese power projection. The end of the Cold War lead to even greater expectations of Japanese rearmament. In 1991, Kenneth Waltz famously argued that even when states do not want to become more militarily assertive, the self-help nature of the international system forces them to convert their economic power into military might. He ultimately claimed that these forces would require Japan to seek great power status and even acquire nuclear weapons. Christopher Layne similarly predicted in 1993 that growing Japanese economic strength would lead it to demand power and become a “much more assertive actor geopolitically.”

These predictions matched the results of a cost-benefit analysis of the buck-passing strategy in the early 1990s. While the benefits of buck-passing remained constant in this era, the risks of the strategy—measured by Japan’s threat environment and its confidence in the U.S. alliance—increased. Tensions in the Japan-U.S. alliance reached a high point in the early 1990s. The alliance was predicated on the fear of a Soviet attack on Japan, a scenario the U.S. desperately wanted to avoid, so the demise of the U.S.S.R. dispelled those fears and led many to question the utility of continued security cooperation. There was a growing fear in Japan that the U.S. would withdraw from Asia to reap its “peace dividend,” and the declining benefits of protecting Japan prompted U.S. officials to re-examine their alliance commitments. In addition to Japan’s fear of abandonment, Japan’s economic ascent generated new friction in its relations with the United States as Americans began to see Japan as a potential threat.

In terms of Japan’s threat environment, scholars still debate whether security risks to Japan increased or decreased during this period. The threat from Russia declined in the short-term, but many strategists were uncertain whether Russia’s economy would recover, which would present a clear hazard to a largely unarmed Japan. Counteracting the moderate decline in Russian threat was a resurgent South Korea, Pyongyang’s emerging nuclear program, and China’s rapidly growing military power. The latter development was particularly impactful given the ongoing territorial disputes between Tokyo and Beijing in the East China Sea. It is hard to say for sure whether these trends produced a net positive or negative impact on Japan’s national security, but even if Japan’s security environment moderately improved, the early 1990s were far
from a time of complacency. It would require a much clearer and more substantial improvement to Japan’s threat environment to overcome its growing doubts in the U.S. alliance.

An uncertain threat environment and intense fears of U.S. abandonment should have pushed a buck-passing Japan towards rearmament. The buck-passing theory would expect Japan to have pursued some combination of two strategies: balancing to reduce its dependence on U.S. assistance and taking measures to minimize the likelihood of U.S. defection. Both these strategies entail an increase in military power. Rearmament undoubtedly furthers the goal of balancing and would have helped Japan prepare for a world where it lacks a firm U.S. commitment. Removing constraints on the SDF would have also proved Japan’s strategic value as a partner by increasing the potential for Japan to support U.S. initiatives. While a full-scale rearmament, especially the pursuit of nuclear weapons, would likely have angered and repelled Washington, the U.S. had long been calling for a larger Japanese military presence to help share the burden of maintaining regional security.

Therefore, although the buck passing theory would expect Japan to resist drastic and costly measures such as nuclearization that would further jeopardize an already-shaken U.S. alliance, it would largely characterize militarization as Japan’s best response to its international position. Three outcomes in the early 1990s would provide evidence that Japan had been following a buck-passing strategy: first, if Japan substantially improved its military capabilities, second, if Japan reduced constraints on the SDF, and third, if it sought to minimize the chance of U.S. defection by demonstrating its strategic value as an ally.

Thus, if Japan were buck-passing, following the conclusion of the Cold War we should have seen Japan substantially improve its military capabilities, reduce constraints on the SDF, and seek to minimize the chance of U.S. defection.

The Empirical Record

During the early 1990s, Japan abandoned its long-standing prohibition on overseas dispatch of the SDF. This was the result of growing pressure for a Japanese human contribution to two campaigns: a U.S.-led war in the Persian Gulf and a UN peacekeeping mission in Cambodia. The U.S. was calling for increased Japanese support for the Gulf War effort, and hawks leveraged these demands to advocate for a reinterpretation of the SDF’s role in international affairs. Likewise, Cambodia had asked Japan to provide military assistance to the peacekeeping effort, which cast doubt on claims that such a deployment would constitute a revival of militarism.

However, the government could not deploy the SDF without explicit Diet approval. Thus, Prime Minister Toshiki Kaifu proposed the 1991 UN Peace Cooperation (UNPC) bill to enable SDF contributions to both missions. Even though the bill did not permit the SDF to employ force, provisions that allowed Japan to cooperate with other countries militaries made it highly controversial. Japanese public opinion firmly opposed the proposed deployments, and opposition groups quickly rallied against the bill. The Japan Socialist Party (JSP) combined forces with an anti-militarist subgroup that defected from the ruling Liberal Democratic Party (LDP) to block the legislation. The UNPC bill’s decisive defeat stood out at a time when political polarization in Japan was incredibly low, with nearly 95 percent of bills passing the Diet. Nor did the bill appear to fail due to strategic concerns. Japan showed no hesitancy to get involved in the Gulf War in other ways, and swiftly committed to non-military measures, including sanctions, humanitarian aid, and financial support to the U.S. forces. Moreover, Diet debate over the UNPC bill seldom discussed Japan’s strategic interests in expanding SDF capabilities, and instead focused on the extent to which Japan ought to remain pacifist.

Despite this initial setback, Japan did manage to secure the SDF’s first ever overseas deployment. After the Gulf War drew to a close in 1991, the Kaifu government used Cabinet orders to circumvent the Diet’s resistance to overseas dispatch. Tokyo deployed its navy to refuel U.S. ships in the Gulf, dispatched the air SDF to ferry refugees, and sent the marine SDF to clear mines. Importantly, however, such actions were only taken after conflict subsided and the Kaifu government justified their legality under the constitution by pointing to their strictly humanitarian aims. Moreover, the symbolic importance of SDF dispatch far exceeded its strategic value: Japanese forces only swept 34 mines out of an estimated 1,200.

Japan sent forces to Cambodia just a year later.
This time, the government could not avoid gaining Diet approval for its actions and introduced the 1992 PKO bill to enable an SDF contribution to the mission. The legislation had to be substantially watered down to appease centrists. It only permitted humanitarian and economic development activities, prohibited the SDF from using force, required that Japan remain neutral, only allowed SDF deployment after a cease-fire, and mandated immediate withdrawal if hostilities broke out. The PKO bill was met with enormous resistance and Pacifists delayed its enactment by five days by delivering their votes at a snail's pace. But in the end, these tactics of last resort could not stop the bill from passing. Japanese forces initially focused exclusively on humanitarian missions such as infrastructure repair, and crucially, the mission never allowed the SDF to employ force. However, the SDF did eventually become involved in some “grey area” operations, such as night patrols.

Outside Japan's contributions to the Gulf War and Cambodian peacekeeping mission, the country showed few signs of remilitarization. An exception may have come in 1995, when the government revised its official military strategy document, the National Defense Program Outline (NDPO) to expand the SDF's mandate from defending against a “small-scale invasion” to allowing a defensive role in “situations in the area surrounding Japan that have a direct impact on the security of Japan.” But this change in mandate still constrained the SDF to solely defensive operations and was accompanied by across-the-board cuts in weapons procurement and force size. And while the SDF shrunk, most East Asian countries ramped up military spending. China in particular—helped along by Japanese aid—underwent a dramatic increase in military power during this period, but Japan refused to improve its power projection capabilities to protect its southern islands from growing Chinese assertiveness.

Besides acquiescing to U.S. demands for a contribution to the Gulf, Japan did remarkably little to shore up the U.S. alliance. Tokyo refused to send ground troops to the Gulf, denied a U.S. request to commit to dispatching troops to the Korean Peninsula in the event of war, and did not make any substantial concessions on the ongoing trade disputes with Washington. The latter was particularly damning as asymmetries in trade and investment were described at the time as the “single biggest irritant” in U.S.-Japan bilateral relations. Thus, while pacifists undoubtedly experienced consequential political defeats during this era, the remilitarization of Japan was far from comprehensive.

Compatibility of the Empirical Record and Hypothesis

Overall, the buck-passing framework struggles to explain Japan's actions during this time period. Although Japan eliminated a central constraint on the SDF, the prohibition on overseas dispatch, it was severely constrained by anti-militarist attitudes. The government could not pass even the highly limited UNPC bill. Instead, the Kaifu administration had to rely on Cabinet orders that limited the SDF's role to post-conflict humanitarian missions. Likewise, to contribute to peacekeeping in Cambodia, the government had to enact less permissive legislation, the PKO bill, to gain the support of centrists in the Diet.

These results may not disprove the buck-passing theory entirely: perhaps the government was motivated to reduce SDF constraints by the strategic considerations analyzed in the previous section. However, it does cast significant doubt on its explanatory power. Regardless of Japan's reasons for increasing SDF autonomy during this period, domestic anti-military attitudes were the most salient feature shaping the outcome of Japan's major foreign policy decisions. Outside the SDF's first and second ever overseas deployments, Japan's military autonomy only increased very slightly during this time period, and such moves did not have clear strategic balancing motives. In fact, Japan ordered force reductions and reduced its military capabilities while China's power grew dramatically. This casts doubt on the realist notion that Japan behaves like any other state, as the buck-passing theory would expect it to enhance its military power when nearby threats arise.

The fear of U.S. abandonment did certainly play a role in Japan's decision to assist the war effort in the Gulf, but one would expect a buck-passing Japan to do a lot more to reassure itself of U.S. commitments. Tokyo's refusal to send troops in the Gulf War seriously hurt its relationship with Washington, and Japan did not respond to any of the U.S.'s other major demands for military action. Such a strained relationship is shockingly dangerous for a buck-passing state fearful
of losing its patron. While Tokyo may not have seen U.S. abandonment as likely, the buck-passing theory would expect it to be highly sensitive to alliance frictions—Japan was decidedly not.

Fundamentally, predictions of remilitarization overlooked the extent of domestic constraints on Japanese militarism,139 and it seems unlikely from the empirical record in the 1990s that Japan follows a strictly rational buck-passing strategy. Table 1 summarizes the key findings.

Conclusions

This paper has argued for the existence of a “buck-passer’s dilemma”: the fear of allied defection should undermine any rational state’s confidence in outsourcing its security to a patron. Japan has often been touted as an instance of the patron-client buck-passing dynamic, but such an explanation fails to account for Japan’s behavior. Although Japan has slowly reduced restrictions on its military, it remains a security outlier, retaining substantial constraints on its own military policy and committing far fewer resources to its armed forces than many would expect. Thus, the buck-passing theory cannot fully explain why Japan has shown a consistent resistance to aggressive military action and has clung to pacifist values despite substantial changes to its threat environment and its faith in U.S. commitments.

The patron-client buck-passing theory’s failure to fully explain Japanese foreign policy indicates a broader flaw in its logic. While scholars should not underestimate states’ tendency to pass the buck in other circumstances, neither should they exaggerate the temptation to free-ride in the security arena: the risks a state faces when fully outsourcing its security are far too great. The buck-passers dilemma represents a major flaw in realists’ explanation for some state’s persistently low military spending, and while questions of burden sharing in alliances will undoubtedly persist, scholars should look to paradigms other than realism to explain why some states neglect their own military development. In evaluating state behavior within alliances, non-realist factors such as domestic politics or norms can provide a fuller picture of the possible outcomes than an approach that focuses solely on international structure.

This result suggests a continuation of the status-quo in East Asia, at least in terms of Japanese policy. Absent an incredibly large and sudden exogenous shock to the U.S. alliance or Japan’s threat environment, Japan will most likely continue its military restraint. Japan may respond to a rising China and growing U.S. isolationism by increasing the autonomy of the SDF and marginally enhancing its capabilities, but it will continue to be seriously constrained by its legacy of anti-militarism. If Washington wants Japan to take on a larger burden in the alliance, perhaps it should slightly back away from its commitments to Japan. Such a move, provided the U.S. does not completely abandon Japan, will pressure Tokyo to take on a larger role in Asian security without prompting it to pursue extreme measures, such as nuclearization or total remilitarization, that may threaten regional stability. Fears of Japanese rearmament are largely overblown, and, for the time being, Japan’s security identity will continue to oppose militarization. Moreover, if one accepts that the buck-passing explanation fails more broadly, the risk that other U.S. allies will rearm in the absence of robust U.S. commitments is likely exaggerated as well. International relations scholars should therefore analyze the specific reasons driving some countries to forsake military power, rather than relying too heavily on overly broad and ultimately flawed structural theories such as buck-passing.

Endnotes


5. I will call this version “patron-client buck-passing” or just “buck-passing.” It is important to distinguish the case in which multiple states of similar strength shirk on their commitments to the alliance from the case of one state becoming wholly reliant on a more powerful ally. The former case, which Mearsheimer (in *The Tragedy of Great Power Politics*, 161) discusses in relation to World War I, can be thought of as a less extreme version of the patron-client buck-passing story critiqued in this paper.


7. I owe Professor Dominic Tierney for his help in coming up with the phrase “buck-passers dilemma.”


16. This is not meant to imply that realists universally find alliances untenable, but instead that realists often argue that states prefer self-reliance when feasible.


19. Realists do acknowledge that the level of mistrust varies between states, and that trust tends to be higher among allies. For example, Joseph Greico models states’ sensitivity to the relative gains problem with a coefficient k that is lower for states in alliances (Joseph Greico, “Anarchy and the Limits of Cooperation: A Realist Critique of the Newest Liberal Institutionalism,” *International Organization* 42, no. 3 (1988): 501).

29. Robert Jervis has argued that the ability to distinguish defensive tactics from offensive ones is a key factor driving the intensity of the security dilemma in “Cooperation Under the Security Dilemma,” *World Politics* 30, no. 2 (1978): 167-214.
32. Lind, “Pacifism or Passing the Buck?” 106.
33. Christensen and Snyder, “Chain Gangs and Passed Bucks,” 145.
41. Lind, “Pacifism or Passing the Buck?” 95, 97, 100.
44. Lind, “Pacifism or Passing the Buck?” 95.
47. Hook, *Militarisation and Demilitarisation*, 76.
49. See, e.g., Lind, “Pacifism or Passing the Buck?” 95.
50. Heginbotham and Samuels, “Mercantile Realism.”
52. Midford, “The Logic of Reassurance.”
53. Kawasaki, “Postclassical Realism.”
56. Kawasaki, “Postclassical Realism.”
57. Lind, “Pacifism or Passing the Buck?” 106.
59. Lind, “Pacifism or Passing the Buck?” 104.
60. Twomey, “Japan, a Circumscribed Balancer,” 186.
71. Oros, Normalizing Japan, 50.
73. Berger, “From Sword to Chrysanthemum,” 134.
75. Berger, “From Sword to Chrysanthemum,” 140.
76. Berger, “From Sword to Chrysanthemum,” 143.
77. Oros, Normalizing Japan, 54.
78. Oros, Normalizing Japan, 10.
80. Hook, Militarisation and Demilitarisation, 119.
82. Oros, Normalizing Japan, 174.
83. Oros, Normalizing Japan, 174.
85. Berger, “From Sword to Chrysanthemum,” 141.
90. Based on SIPRI’s military expenditure database.
93. Oros, Normalizing Japan, 63.
94. Oros, Normalizing Japan, 68.
96. Hook, Militarisation and Demilitarisation, 118.
98. Hook, Militarisation and Demilitarisation, 115.
99. Lind, “Pacifism or Passing the Buck?” 120.
106. For works analyzing the post-Cold War Era as a turning point in Japanese foreign policy, see Hook, Militarisation and Demilitarisation; Oros, Normalizing Japan; Midford, “The Logic of Reassurance;”Samuels, “‘New Fighting Power!’”
111. Heginbotham and Samuels, “Mercantile Realism,” 179.
113. Lind, “Pacifism or Passing the Buck?” 110.
116. Oros, Normalizing Japan, 44.
117. Hook, Militarisation and Demilitarisation, 77-79.
118. Hook, Militarisation and Demilitarisation, 91.
120. Hook, Militarisation and Demilitarisation, 109-112.
122. Hook, Militarisation and Demilitarisation, 83.
123. Hook, Militarisation and Demilitarisation, 80.
124. Hook, Militarisation and Demilitarisation, 83
125. Katzenstein and Okawara, “Japan’s National Security,” 113
126. Oros, Normalizing Japan, 86.
127. Hook, Militarisation and Demilitarisation, 86.
128. Hook, Militarisation and Demilitarisation, 87.
131. Hook, Militarisation and Demilitarisation, 89.
132. Hook, Militarisation and Demilitarisation, 93.
133. Hook, Militarisation and Demilitarisation, 87.
140. Oros, Normalizing Japan, 72.

Bibliography


tution/.


institute.org/the-interpreter/korean-war-ii-maybe-not-likely.


Tribal Affiliation or Class-Based Differentiation? Class, Tribal Identity, and Interpretations of State-Led Tourism Development in Petra, Jordan

Nicolas Reeves, George Washington University

Introduction

Shortly after visitors enter Petra Archaeological Park, small groups of children approach them to sell them rolls of postcards that depict the Treasury, the Monastery, and the city’s other architectural marvels. Passing these initial elements of Petra’s tourism economy and upon reaching the Treasury, tourists are quickly confronted by a wave of jewelry salesmen, tour guides, and postcard sellers. This commercial activity accompanies visitors as they venture further into the park, where tent-like souvenir stands dot the landscape and men on horse-, donkey-, or camelback encircle visitors to offer them a ride. Consequently, even a casual tourist quickly realizes the degree to which the livelihoods of members of Petra’s local Bidūl and Layāthnā tribes depend on tourism-related work.

Petra welcomed 828,952 visitors in 2018, making it Jordan’s most popular tourist destination.1 The city, however, was a scarcely visited rural backwater as recently as 50 years ago. During this time, only a minority of Petra’s tribal inhabitants worked in tourism; in fact, many of the region’s indigenous Bidūl and Layāthnā tribespeople raised livestock, worked in agriculture, and used the city’s monuments and caves as winter homes. Remembering this period, older community members reminisce about the time “when people depended on the land and each other.”2 As government efforts to turn Petra into a key source of tourism revenue intensified during the 1980s, the lives of the region’s native inhabitants changed significantly. In 1984, state authorities relocated the Bidūl from Petra’s caves to government-built housing units in the newly-built village of Umm Sayhoun. Meanwhile, the Layāthnā began building tourist-oriented restaurants and hotels in their hometown of Wadi Musa to serve the rising number of visitors coming to witness Petra’s architectural marvels.

Tourism had transformed life in Petra completely by 1995. Most Bidūl and Layathna families had become financially dependent on work in the tourism sector, which remains the case to this day. Moreover, Petra’s 1985 selection as a UNESCO World Heritage Site and its 2007 designation as one of the New Seven Wonders of the World have driven investment in the region by multinational hotel chains, including Marriott and Movenpick. Combined with state efforts to use profits from Petra to benefit Jordan’s ailing national treasury, these features of the city’s development over the last 40 years have led to the creation of a complex web of tribal, governmental, and international stakeholders that compete for access to tourism revenues in the ancient Nabataean capital.

Against the backdrop of Petra’s ongoing commercialization, I interviewed 56 Bidūl and Layāthnā tribespeople between August 2018 and March 2019 to
determine the influence of tribal identity as a differentiating factor between narratives surrounding local access to the city’s tourism economy. Using positioning theory, I found that both class and tribal differences play a role in explaining the distinctions between local discourses related to competition over Petra’s tourism revenues. Specifically, tribal affiliation constitutes the primary distinguishing factor between oral histories propagated by study participants; these histories demonstrate their Bidūl or Layāthnā forefathers’ special connection to Petra, and claim their tribe’s right to a larger share of Petra’s tourism-related financial benefits than the region’s other indigenous groups. On the other hand, in narratives related to intervention in local affairs by regional and national authorities, class played a larger role than tribal identity in distinguishing between positioning statements propagated by Bidūl and Layāthnā interviewees. The salience of class in interviewees’ perceptions of government politics towards Petra carries significant implications for policymakers’ efforts to accommodate indigenous interests in tourism development plans, particularly given that policies designed to foster local community growth target individual tribes, not specific classes. Government officials must take steps to involve tribespeople from all elements of the city’s socioeconomic spectrum in development policymaking in order to limit the danger of excluding underrepresented groups from this process.

**Literature Review**

The local impacts of the struggle over access to Petra’s tourism revenues manifest in the ways in which members of the Bidūl and the Layāthnā position themselves, their tribe, and other tribal and governmental stakeholders in the city’s tourism economy. In addition, this dispute is shaped by historical tribe-state relations in the Petra region and by the dynamics of social and economic advancement in Jordanian society more broadly. In this literature review, I describe the tenets and origins of positioning theory and how this framework applies to the study of intergroup conflict. I then employ these concepts to provide an overview of Jordan’s socioeconomic class structure. Finally, I use positioning theory to describe the interaction between state and tribal institutions in the Petra region from 1967 to the present.

**Positioning Theory: Origins, Development, and Basic Principles**

The framework of positions and positioning represents a dynamic alternative to the idea of role in discourse analysis. In contrast to the static notion of role, a position is a discursively negotiated set of rights, duties, and obligations held by a participant in a local moral order at a given moment in time. Consequently, positioning constitutes the act by which a speaker assigns a position to another individual. Positioning theory analysis, then, is the description of the dynamic process through which interlocutors negotiate and renegotiate the arrays of rights, duties, and obligations assigned to them by other members of a local moral order.

In the social sciences, this conceptualization of positioning was first employed in Wendy Hollway’s (1984) analysis of the construction of subjectivity in heterosexual relations. Rom Harré (1990) applied the idea that positions are negotiated on a context-specific basis to the discursive formation of identity. Harré and Luk van Langenhove (1991) then used positioning theory analysis to examine interpersonal disputes, after which Sui-Lan Tan and Fathali Moghaddam (1999) expanded positioning theory to the analysis of intergroup conflict. Since then, the work of positioning theorists such as Tobias Greiff has helped inspire a paradigm shift in conflict resolution by challenging the field’s traditional emphasis on resolving violent disputes. The positioning theory approach to examining conflict instead shifts the analytical focus towards exploring discursive processes that may lead to future violence among members of a local moral order. This approach has been employed by Tom Bartlett and Marshaley Baquiano to examine the narrative context of disputes between indigenous communities and state entities in Guyana and the Philippines. My study uses positioning theory to analyze conflict between tribes, regional authorities, and the national government in Jordan.
Figure 1: The Speech Act - Position - Storyline triad. Source: Harré and van Langenhove, “Introducing Positioning Theory,” 18.

In positioning analysis, the idea that conflict does not occur in a contextual vacuum is captured in the interaction between positioning statements and storylines. As Figure 1 shows, positioning acts generate meaning against the complex backdrop of a storyline, enabling a person's actions to become intelligible and relatively determinate within a local moral order, which is defined as the negotiated discursive structures and practices of participants in a group.¹⁰ Since storylines themselves consist of groups of positioning statements pertaining to a common narrative, it follows that a positioning act's significance is determined by the extent to which it is incorporated into storylines that are salient in a given moral context. In practice, this occurs when other members of the local moral order continue to interact with previously-propagated positioning statements.¹¹ This concept is embodied by second- and third-order positioning acts, which are speech acts in which a positioned interlocutor rejects a speaker's original, first-order positioning act.¹² For instance, Alia, a hypothetical souvenir seller in Petra, would engage in second-order positioning by refusing a police officer's order to dismantle and remove her shop in Petra. She would be engaging in third-order positioning if, after dismantling the stand, she conversed with her friend Mona and questioned the police officer's jurisdiction to order her to do so. These examples show that the authority of positioning acts in a storyline can be questioned at any time. As storylines develop, actors renegotiate their rights, duties, and obligations as their ability to do so within the constraints of the local moral order evolves.¹³

Positioning Theory, Intergroup Relations, and Conflict

As discussed earlier, the positioning theory approach to conflict involves descriptively analyzing the discourses that inform group conflict on a case-by-case basis.¹⁴ Positioning theory thus enables the identification of undesirable storylines and positions that could endanger actors' agency before they erupt into violence.¹⁵ In his study of the politicization of public life in Bosnia and Herzegovina after the signing of the Dayton Peace Accords in 1995, Greiff emphasized the relevance of forced and frozen positions in this context. According to Greiff, individuals forcibly position other agents by coercing them to occupy a position that differs from their chosen position. Individuals occupy a frozen position when they are forced to adopt a cluster of rights, duties, and obligations that do not permit them to discursively reposition themselves within the constraints of the local moral order.¹⁶ Since violence frequently follows such acts that limit individual or group agency, positioning theory analysis can reduce the danger of a conflict escalating by identifying situations in which frozen positions might develop.

The examination of pre-positioning and group memory plays a crucial role in identifying situations where forced and frozen positioning could occur. In his study of the role of time and context in positioning theory, Bartlett found that certain memories stand out more than others in the development of storylines that impact the local moral order in which a group participates in the positioning process.¹⁷ In Guyana, for example, the Amerindian Act of 1976, which established centralized control over the country's indigenous Amerindians, created a fissure within Amerindian history against which all subsequent events have been measured.¹⁸ Beyond figuring prominently in the daily lives of Amerindians, the Act, along with documents written in response to it, like the 2001 Constitution of the North Rupununi District Development Board, constitutes a portion of the national-level storyline that describes negotiations between Amerindians and the government over the status of the Amerindians.¹⁹ By becoming essential pieces of the Amerindians' shared memory, these positioning acts transcended their original spatial and historical context, enabling their ongoing impact on the conflict's development.

In a similar way, Andrew Shryock's (1997) study of oral histories related by ‘Adwani and ‘Abbadi tribespeople of Jordan's Balga region demonstrated the ongoing role that shared tribal memories play in informing relations between the two groups in the
present. Specifically, Shryock observed that oral accounts of the past focusing on tribal affairs in the Balga before the advent of Hashemite rule in Jordan rejected the Western historiographical tenets of textuality and temporality; this allowed for continuous revision and re-appropriation of these accounts to fit the contexts in which they were told. By recognizing that the Balga's current tribal inhabitants view the region's pre-1921 history as practical, context-dependent, and in a constant state of reconstruction, Shryock revealed that 'Adwanis' and 'Abbadi' reproductions of pertinent events in their tribes' pasts can be analyzed for narrative clues about past and present power dynamics between these two groups.

Shryock's examination of the contextual, content-related, and performative aspects of Balga tribespeople's oral accounts of the past led him to identify 'Adwani historical narratives as a “dominant discourse” and 'Abbadi varieties as a “subaltern” one. For example, 'Adwani tribesmen adopted an authoritative tone as they related their tribe's most significant historical accomplishments to Shryock. 'Adwani narrators also insisted that Shryock mention them by name in his study, thus discursively linking their identities to the illustrious feats of former 'Adwani sheikhs to whom they were related. 'Abbadi tribesmen, in contrast, preferred to remain anonymous, telling stories about the tribe's collective achievements that contained only sparing individual references. Moreover, 'Abbadi historical self-positioning practices always entailed placing the tribe in opposition to the 'Adwani. 'Adwani historical narratives, however, rarely mentioned the 'Abbadi. In fact, 'Adwanis asserted that the 'Abbadi - 'Adwani war, the central event of 'Abbadi oral histories, did not even occur. For Shryock, the disparities between orally-produced 'Adwani and 'Abbadi accounts of the Balga's tribal past revealed the uneven power dynamics that characterize the two groups' relations, both historically and in the present. The contextual connection between my Bidūl and Layāthnā participants' attempts to justify their right to benefit from tourism in Petra and their accounts of the city's tribal past reveals that Shryock's insights are applicable to the study of the power dynamics between tribal stakeholders in Petra's tourism economy today.

Positioning, Class, and Social Mobility in Jordan

On the individual and the tribal level, positioning plays an important role in determining class differences in Jordan. Because large portions of the material resources available to Jordanians are distributed through political channels—whether via the state or the many parallel donor-funded agencies that also provide public services—social mobility is often determined by a person's personal connections and ability to leverage them, a reality embodied by the concept of wāsita. Aseel Al-Ramahi described wāsita as a form of intercession in which individuals leverage personal relationships with politically powerful people to obtain goods, services, or favors they otherwise would not be able to receive. Wāsita-based connections are particularly important for individuals seeking benefits from the government, such as jobs, scholarships, or professional licenses. For instance, nearly half of respondents to a 2000 survey of 320 politically active Jordanians reported that they would first seek employment through wāsita before searching for jobs through formal channels. Another poll conducted in 2005 revealed that 86 percent of surveyed Jordanian businesspeople consider wāsita important for doing business in the public sector, while 55 percent admitted to using wāsita themselves. These statistics demonstrate that in the public sector in particular, wāsita is entrenched to the point that most citizens view it as a necessary means for socioeconomic advancement.

Wāsita's prevalence in public sector political, economic, and social interactions enables Jordanians from East Bank tribal backgrounds—in contrast to the country's majority population of citizens of Palestinian origin—to dominate high-ranking positions in state entities ranging from the military to government ministries. This is because the leaders of Jordanian tribes from East Bank origins, such as the Banī Sakhr, Banī Hassan, or the Bidūl and Layāthnā in Petra, represent the key intercessors through whom wāsita translates into political and social capital. Therefore, an individual's position and positioning relative to an influential intercessor from his/her tribe constitute powerful determinants of class differences in Jordan.

The sheikh, or chief, represents the paramount tribal intercessor. Sheikhly power is recognized by the king himself, as the name of each Jordanian sheikh is written down at āl-Dīwān āl-Malakī [The Royal
Hashemite Court]. Beyond legitimizing the sheikhs’ authority, āl-Dīwān provides a forum for these influential representatives of the monarch’s key support base to voice complaints or lobby on their behalf. While in theory, sheikhly access to āl-Dīwān has improved the representation of tribal interests in Amman, several members of the Layāthnā tribe pointed out that the opposite has occurred. Participant 29, a taxi driver, revealed that “sheikhs are no longer responsible to the tribes they represent. Before, it was possible to get rid of a bad sheikh through popular consent. Now, with approval from al-Dīwān al-Malakī saying that ‘you are the sheikh,’ this responsibility to the people disappears…The sheikhs take the government salary from al-Dīwān al-Malakī and do not do anything good for the people. They extract rents from all the places in the local community, like the mosques.” Participant 29’s assertion reveals that instead of reinforcing the role of sheikhly intercession, al-Dīwān has caused the link of accountability between sheikhs and their tribal constituents to disappear, indicating that this institution may have actually reduced ordinary tribespeople’s access to wāsita.

In addition to al-Dīwān al-Malakī, John Shoup (1990) revealed that the re-establishment of the Ottoman-era position of mukhtār [headman] in the 1950s has also impacted sheikhly authority and the means through which Bidūl and Layāthnā tribespeople access wāsita. The mukhtār, though he is an elected government official, has similar responsibilities to those of the sheikh, causing the authority of the mukhtār to sometimes rival the sheikhs. This is the case with the Bidūl; during his 1983-1988 fieldwork among the Bidūl, Shoup noted that Umm Sayhoun’s mukhtār always convened the tribe’s majlis (council) at his home in the village, a duty traditionally performed by the sheikhs. Moreover, he had taken over the job of lobbying the state for goods and services on behalf of the tribe. While Wadi Musa’s mukhtār has not encroached on the traditional duties of the sheikhs of the Layāthnā to this degree, he nonetheless told me during my 2018–2019 fieldwork that his responsibilities include coordinating with local government officials to ensure that development policy initiatives heed the interests of Layāthnā tribespeople. In the Petra region as a whole, then, the establishment of al-Dīwān al-Malakī and the creation of the position of mukhtār have contributed to both the expansion and contraction of avenues of tribe-based wāsita available to members of the Bidūl and Layāthnā.

Historical Background: Local Impacts of Tourism Development Efforts in Petra

In addition to the evolution of the concept of tribal wāsita in Petra, the ongoing impacts of state development policies in the region over the last 50 years have influenced the current conflict over tourism resources between the Bidūl, the Layāthnā, and Petra’s local government – the Petra Development and Tourism Region Authority (PDTRA). In this section, I apply the tenets of group positioning theory to the history of state-led tourism development in Petra since the 1967 Arab-Israeli War, called āl-Naksa (‘The Setback’) in Jordan. This overview will trace the development of the diverse positions held by members of the Bidūl and Layāthnā tribes within the storyline of economic competition in Petra’s tourism economy.

Twenty years after annexing the West Bank following the 1948 Arab-Israeli War, Jordan lost control of the territory following the war in 1967. As Salam Al-Mahadin argued in her 2007 critical analysis of Petra’s emergence as Jordan’s tourism capital, ceding the West Bank represented a naksa for Jordan and its king at the time, Hussein bin Talal, on multiple fronts. First, the retreat from Palestine eliminated what little legitimacy the monarch had to position his country as the embodiment of the pan-Arab dream of Sharif Hussein bin Ali, Hussein’s grandfather, and the inspirational force behind the 1916–1918 Arab Revolt. Furthermore, the loss of the symbolic and economically important city of Jerusalem severed Jordan from its most important tourism revenue source and curbed the country’s ability to position itself internationally as “The Holy Land.” While āl-Naksa immediately constrained Jordan’s position both economically and politically, it took three years for these outcomes to fully manifest in the form of dramatic national repositioning efforts undertaken by the monarchy in response to the events of September 1970, or Black September.

Beyond its effect on Jordan, āl-Naksa was a disaster for Palestinians living in Palestine and elsewhere in the Arab world. In particular, losses of Egyptian, Jordanian, and Syrian-held Palestinian territories in Gaza, Jerusalem, the West Bank, and the Golan Heights galvanized the Palestinian nationalist move-
ment, which drew strength from refugee and diaspora populations in several neighboring states—most notably Jordan. The increasing assertiveness and militancy of the Palestine Liberation Organization (PLO) paved the way for fighting between the PLO fidā‘iyīn (Palestinian freedom fighters) and forces loyal to the Hashemite monarchy in 1970 in a series of violent confrontations that became known as Black September. The expulsion of fidā‘iyīn leaders to Lebanon after Black September was mirrored in King Hussein’s purge of citizens of Palestinian descent from the state’s bureaucracy and military and their replacement with “ethnic Jordanians” that represented tribes from regions east of the Jordan River. King Hussein’s efforts to “Jordanize” the government and security apparatus became emblematic of the monarchy’s positioning efforts to create a new national storyline in which East Bank Bedouins—Jordan’s original inhabitants, in contrast to the majority population of Palestinian migrants and refugees—became the ultimate symbol of Jordanian identity. Bedouin traditions, meanwhile, became the essential elements of Jordan’s national heritage.

As was the case with the East Bank and its ethnic inhabitants as a whole, the symbols of Petra’s glorious Nabataean past became the focal point of government efforts to reposition Jordan as a country whose national fabric developed around the customs and traditions of the Bedouins. Petra lent itself to this endeavor due to the ease with which state narratives could link the area’s rich Nabataean history to its present-day tribal inhabitants, who lived in close proximity to the city’s imposing temples. Layān Fuleihan argued that because they lived in the city’s caves, the Bidūl represented a physical and spatial link to Petra which gave King Hussein the symbolic currency necessary to root Jordan’s discursively-produced tribal origins in the civilization of the Nabataeans. The manufactured Nabataean relationship was further cemented when Petra’s Bedouins were inscribed on the United Nations Educational, Scientific and Cultural Organization (UNESCO)’s intangible heritage list in 2005, giving international recognition to the government’s positioning efforts. This legitimization papered over the primary flaw of the Hashemites’ Nabataean heritage narrative: the Bidūl tribe’s sedentarized existence in Petra’s caves for hundreds of years conflicted with their post-Black September status as key carriers of Jordan’s nomadic Bedouin image. Furthermore, the government’s decision to resettle the Bidūl in state-sponsored housing in nearby Umm Sayhoun to protect the city’s archaeology and create a less hectic environment for western tourists after UNESCO’s 1985 recognition of Petra as a World Heritage Site created further inconsistency in this regard. To this day, these contradictions between the Bidūl’s sedentary, modern lifestyle in Umm Sayhoun and their forced positioning as representatives of Jordan’s traditional, nomadic past is used by Layānthnā and PDTRA stakeholders to cast doubt on the legitimacy of the tribe’s participation in Petra’s globalized tourism economy.

While the Bidūl were positioned as messengers from Jordan’s Nabataean past by the politics of place, the Layānthnā became the chief local economic beneficiaries of government-led regional tourism development efforts through Wadi Musa’s strategic location at Petra’s entrance. Although the building of a government rest house halfway between Wadi Musa and the Treasury in 1950 represented the extent of Amman’s interest in developing Petra for tourism until the 1980s, John Shoup (1985) observed that in fact, this single structure placed the Layānthnā in a prime position to participate in the tourism economy at the expense of the Bidūl. The rest house’s strategic location for the Layānthnā allowed them to sell postcards, tour books, stamps, souvenirs, and horse rides only a short walk from their homes, enabling the tribe to establish an early monopoly over the legal tourism hospitality...
Thirty years later, Layāthnā entrepreneur Hani Fallahat contracted with Jordan’s national bus company – Jordan Express Tourist Transport (JETT) – to bring its service to Petra. This deal brought about an additional economic windfall for the Layāthnā. Fallahat became JETT’s official tour guide, making sure to bring visitors to Layāthnā-owned souvenir stands on their way around Petra. Fallahat’s bargaining also ensured that JETT buses would arrive at Wadi Musa’s lower parking lot, a move that created a financial boon for the city’s hotels, whose total capacity expanded from 55 beds in 1980 to 3,682 in 1998. Wadi Musa’s restaurants and hotels thus became the focal point of the lucrative hospitality business of Petra’s tourism industry. To this day, the Bidūl, who live four kilometers away in Umm Sayhoun, do not benefit from this aspect of the tourism sector or its spillover effects, such as the provision of basic infrastructure including pharmacies, 24-hour health centers, and ease of access to public transportation connecting the community to the rest of Jordan.

While tourism has transformed the Petra region from a desert backwater into an urban hub, tourism development has proceeded unevenly, creating vast economic gaps not just between the residents of Wadi Musa and Umm Sayhoun, but within the Layāthnā tribe as well. In Wadi Musa, top-down efforts to develop the hospitality industry of the tourism sector froze many poor Layāthnā out of this lucrative business. The desire to attract affluent tourists to Petra led the government to offer incentives for building four and five-star hotels in the region. Since most locals cannot afford to build such establishments, these businesses are typically owned by foreign investors or wealthy elites from Amman. According to economists Nasim Barham and Horst Kopp, the state’s focus on financing high-end luxury hotels, combined with its overly optimistic growth prognoses, have resulted in over-building in the Petra region. These policies have led to a chronic surplus of hotel beds in Wadi Musa, instigating a race to the bottom for hotel prices as the owners of these establishments attempt to attract customers. This trend cripples local owners of non-luxury hotels, who possess neither the savings nor the access to the government financing needed to run their operations at a loss for extended time periods. Furthermore, although Layāthnā-owned souvenir shops, in contrast to most of those run by the Bidūl, have government permits, these establishments are subject to high rents paid monthly to the Ministry of Tourism and Antiquities, leaving few profits with which to support their families.

The hardships the Bidūl and Layāthnā face in the midst of top-down government efforts to attract wealthy tourists to Petra reflect a growing trend across the country that contradicts the monarchy’s positioning of the Bedouin as post-Black September Jordan’s paramount symbol. Since King Abdallah II came to power in 1999, regional wars and chronic fiscal crises have forced him to initiate policies to liberalize the Jordanian economy to preserve the country’s access to International Monetary Fund loans and donor funds. As a result, the Bedouins’ privileged position in state discourse on national identity, propagated both within Jordan and abroad, contrasts the harsh economic realities of ordinary East Bank tribespeople, who have suffered financially from King Abdallah’s efforts to reduce the size of the country’s public sector, the main employer of “ethnic Jordanians” following the events of Black September. In Petra, the king’s decision to embrace the neoliberal economic policies championed by the Washington Consensus resulted in the establishment of the PDTRA itself in 2009. Since its creation, the PDTRA has developed new incentives for international investors, including unlimited repatriation of profits and simplified procedures for employing non-Jordanians in foreign-owned businesses. At the same time, members of both tribes complain of a lack of opportunities for local community members in this government body.

It is against this background of transformation that the participants in my study position themselves, their tribe, and other tribal and governmental stakeholders in Petra’s tourism economy. Much in the way that Bartlett and Shryock observed the continued impact of seminal positioning acts on Amerindians in Guyana and ‘Adwani and ‘Abbadi tribespeople in Jordan’s Balga region, the social force of past government policies towards Petra – particularly the 1986 resettlement of the Bidūl and the creation of the PDTRA – manifests itself in their prominence in positioning statements propagated by the Bidūl and the Layāthnā today. Positioning acts and pertinent episodes of pre-positioning thus create storylines through which the stakeholder conflict over Petra’s limited tourism business.
resources can be analyzed.

**Methodology**

This study employs a participant-centered approach to explore how economic hardship and governmental development policies impact the discursive strategies employed by the Bidūl and Layāthnā to position themselves and competing stakeholders in Petra's tourism economy. This choice of methodology centers around the recognition that study participants, as active participants in Petra's local moral order, represent the true experts in this inquiry. This idea is reflected in the semi-structured interview method employed in this study. The questions below were designed to serve as simple conversation starters, enabling participants to quickly take control of the discussion and choose the narratives of local life they wished to convey:

1. How does your life differ from that of your father and grandparents?
2. What is your tribe’s history in Petra?
3. How has tourism impacted your life? How has this impact changed across generations?
4. Has tourism impacted your relationship with other governmental or tribal stakeholders?
5. If you could change one thing about your current situation, what would it be?

During interviews, study participants re-described past interactions and events related to these five questions that involved their tribes and other members of Petra’s moral order. The author’s role as an interlocutor in these discussions remained limited to that of a listener, with the exception of posing follow-up questions in order to contextualize historical events, such as battles during Ottoman times, that study participants claimed were instrumental to understanding Petra’s local moral order today.

Although I tried to limit the influence that preconceptions about Petra’s local community would have on my analysis of positioning statements propagated by study participants, time constraints forced me to make informed assumptions about the tourism economy to guide my purposive sampling method. Specifically, I assumed that if tribal identity was a differentiating feature among discourses surrounding competition in Petra’s tourism economy, this distinction would be more conspicuous among the Bidūl and the Layāthnā than among the region’s other tribes, like the ’Ammārin. Shoup’s 1983-1988 fieldwork visits to the Petra region indicate that this assumption is reasonable, justifying the decision to limit the study sample to members of the Bidūl and Layāthnā only. While the construction of the rest house, the bus station, and Umm Sayhoun village precipitated the Bidūl and Layāthnā’s present reliance on work in the tourism sector, Shoup (1990) observed that the preference of the ’Ammārin for nomadism and the tribe’s location in Bayda kept the group on the tourism economy’s periphery.52 Bille (2019) also revealed that ’Ammārin tourism entrepreneurs work primarily as souvenir sellers, guides, and camp hosts at the Little Petra archaeological site near Bayda, an observation that was confirmed by two study participants during my fieldwork.53 Moreover, interviewees’ positioning statements surrounding the intertribal dimension of the stakeholder conflict over access to Petra’s tourism economy focused exclusively on competition between the Bidūl and the Layāthnā.

Between August 2018 and March 2019, I conducted fifty-six face-to-face, semi-structured interviews with Bidūl and Layāthnā tribespeople in Wadi Musa, Petra, and Umm Sayhoun.54 I conducted the interviews in Arabic. Twenty-eight interviewees were from the Bidūl tribe, while the other twenty-eight were from the Layāthnā. I coded the subjects such that Participants one through twenty-eight are interviewees from the Bidūl and Participants twenty-nine through fifty-six are from the Layāthnā. My first interviewees were souvenir sellers in Wadi Musa and Petra Archaeological Park. I found additional participants through snowball sampling. Specifically, I asked interviewees to introduce me to other local community members who they perceived as belonging to a higher socioeconomic class. This strategy allowed me to interview tribespeople from across Petra’s socioeconomic spectrum and enabled me to learn how individuals in both tribes perceive their own standing within Petra’s class structure.

**Classes in Petra**

Bidūl and Layāthnā tribespeople’s positioning statements indicated that Petra’s local community can be divided into an entrepreneurial, bureaucratic, and ruling class, thus mirroring James Bill’s argument regarding the socioeconomic structure of tribal and
monarchic societies like Jordan’s. In Petra, these socioeconomic groups differ from one another based on how their members earn income and based on wāsita’s role in facilitating entry into the class. With respect to the difference between the bureaucratic and the entrepreneurial class, participants from both tribes explained that in contrast to entrepreneurs, whose incomes depend on the quantity of a good or service they sell, bureaucrats receive a fixed monthly salary. Furthermore, a significant portion of Bidūl and Layāthnā interviewees stated that access to wāsita constitutes a major factor in obtaining the government positions that make up the bureaucratic class. Finally, the ruling class, in contrast to the entrepreneurial and bureaucratic class, is connected with powerful actors within Jordan’s state apparatus, enabling members of the class to intercede on behalf of other participants in Petra’s moral order in wāsita-based relationships. Next, I describe the Bidūl and Layāthnā tribespeople that make up each class and the dynamics of advancement within each of these groups.

The Entrepreneurial Class

Forty-four of the participants in my study are from the entrepreneurial class. Twenty-eight of these are souvenir sellers, two are donkey owners, two are horse owners, two are refreshment stand owners, one is a tour guide, one is a taxi driver, seven are hotel managers, and one is a restaurant owner. Within the entrepreneurial class, Bidūl and Layāthnā interviewees made conscious distinctions between shughl qānūnī [legal work] and shughl mukhālaf [illegal work]. Positioning statements differentiating shughl qānūnī from shughl mukhālaf were salient particularly among souvenir sellers, donkey owners, and horse owners from both tribes, who used this dichotomy to criticize the flaws of the permitting system that governs the participation of tribal entrepreneurs in the tourism economy. For instance, Layāthnā souvenir sellers and horse owners employed this classification to criticize the flaws of the permitting system that governs the participation of tribal entrepreneurs in the tourism economy. For instance, Layāthnā souvenir sellers and horse owners employed this classification to criticize the flaws of the permitting system that governs the participation of tribal entrepreneurs in the tourism economy. For instance, Layāthnā souvenir sellers and horse owners employed this classification to criticize the flaws of the permitting system that governs the participation of tribal entrepreneurs in the tourism economy.

Beyond Participant 15, the majority of the other twenty illegal souvenir sellers that participated in this study also argued that the classification of work as simply mukhālaf or qānūnī ignores the sociopolitical realities that render obtaining a permit a near-impossible task for Bidūl tribespeople in particular. In this regard, one mukhālaf souvenir shop owner from the Bidūl asserted that “getting a permit [for a legal souvenir shop] requires wāsita.” This statement demonstrates wāsita’s salience with respect to socioeconomic advancement within the entrepreneurial class. It is important to note as well, however, that interactions between tribal entrepreneurs and tourists that I observed over the course of my fieldwork in Petra reveals that success within this portion of the tourism economy also depends on actors’ advertising skills and the quality of service they provide to customers. For instance, qānūnī and mukhālaf souvenir sellers recite slogans in multiple languages to convince tourists to consider buying items on display, a practice that is repeated by restaurant managers as tourists return to Wadi Musa in the evening.

The Bureaucratic Class

I interviewed four people from the bureaucratic class. Of those, three work for the Capital of the Nabataeans Cooperative Society, a government-affiliated local community development organization, and the other is a manager in the PDTRA. Wāsita constitutes the
primary means for social advancement within the bureaucratic class. Particularly in the case of jobs at the PDTRA, familial connections to management-level PDTRA officials can almost be considered a requirement for employment. For instance, Participant 29 said, “there are plenty of qualified people with university degrees who are waiting on the list to work at the PDTRA, but people with wāsita just get in. They can apply and start working the next day.” Figure 2 illustrates this difference; the borders of the bureaucratic class reach higher than those of the entrepreneurial class, portraying the superior levels of wāsita bureaucrats can obtain in comparison to their entrepreneur counterparts. In the class diagram, the bureaucratic class shares a border with the ruling class, revealing that the bureaucratic class serves as a conduit to the highest positions within the PDTRA, which are part of the ruling class.

Figure 2: Displays Petra’s tribal class structure and the occupations associated with each class. The number of people interviewed is written in parentheses next to each form of employment.

**The Ruling Class**

I interviewed eight members of Petra’s ruling class. Four of these are sheikhs from the Layāthnā, two are sheikhs from the Bidūl, and one is an advisor to the Deputy Chairman of the PDTRA Board of Commissioners. I also interviewed Wadi Musa’s mukhtār. The ruling class is simultaneously static and dynamic, mirroring the characteristics of national state institutions in Amman in this respect. On the one hand, the tribal chieftaincy is passed between generations on a hereditary basis, meaning that the portion of the ruling class encompassed by the sheikhs is characterized by hereditary rigidity, similar to the institution of the royal family. In contrast, frequent turnover exists in the position of mukhtār and in the leadership of the PDTRA, as is the case with parliamentary leadership positions in Amman.

Figure 3: The salience of class and tribal identity in the two identified themes.

**Positioning Theory and Theme Analysis**

By conducting a positioning theory analysis of the fifty-six interviews, I identified two themes that resonate across storylines propagated by my Bidūl and Layāthnā participants. In storylines associated with “Theme 1: Tribal Narratives,” members of each group use competing versions of Petra’s tribal history to justify why either the Bidūl or the Layāthnā deserve to receive a larger portion of the city’s tourism-linked financial benefits today. In discourse surrounding “Theme 2: Outside Intervention” in Petra’s development, however, the intratribal agreement that characterizes positioning strategies employed in participants’ tribal narratives breaks down. Instead, this theme features primarily class-differentiated interpretations of local and national government tourism development policies towards Petra, demonstrating that class affiliation represents the key differentiating factor among opinions of Bidūl and Layāthnā tribespeople with regard to the politics of the PDTRA and state authorities in the Petra region.
Tribal Affiliation of Class-Based Differentiation?

Discussion

Tribal narratives and local storylines related to outside intervention in Petra's development shed light on the challenges faced by members of Petra's local community as they compete in the city's tourism economy. In this section, I use positioning theory to explore the nuances of the tribe- and class-based lenses through which my interviewees conveyed their group histories, their socioeconomic position in Petra's local moral order, and the obstacles complicating their advancement within it.

Tribal Narratives

Bidūl and Layāthnā participants mobilized their tribal identities in positioning statements aimed at critiquing the current distribution of opportunities for social advancement between the two groups today. To justify these criticisms, interviewees recounted competing versions of Petra's tribal past that highlighted their Bidūl or Layāthnā forefathers' connection to the city while simultaneously casting doubt on the other tribe's historical association with Petra. These tribal narratives thus attempt to legitimize one tribe's current presence in Petra at the expense of the other, creating an implicit rationale for why either the Bidūl or the Layāthnā deserve a larger proportion of the economic benefits associated with visitor flows to Petra in the present.

In contrast to the Layāthnā, who predominantly live and work outside of Petra Archaeological Park, the Bidūl's continued presence in and around the ancient Nabataean city's famous structures supports the tribe's self-positioning as a group whose historical connection to Petra is unrivalled. In fact, four mukhālaf Bidūl souvenir sellers asserted that “the Nabataeans are [our] ancestors,” solidifying this geographic link to the city and its ancient builders. Although this idea did not feature explicitly in the historical narratives of other Bidūl interviewees, Participant 22 captured the generally-held consensus in her tribe by asserting, “The Bidūl are a group that is present in Petra and in āl-Hamayma, a region located on the way to Aqaba near Wadi Rum...Three-hundred years ago, all of them were in āl-Hamayma, but then two sub-tribes of the Bidūl, the Samāhīn and the Faqīr, departed and went to Petra through the valley.”

Beyond positioning themselves as longtime inhabitants of Petra, Bidūl interviewees also claimed that the Layāthnā lived far away from the city in the past. Participant 7 recounted, “The Layāthnā were in āl-Hay. We were the only ones here. There were twenty families, but by the grace of God, people began to marry, and the population increased.” By positioning the Layāthnā as the historical inhabitants of “āl-Hay,” a stretch of land located on the flat plateau that extends to the North beyond the hills of Wadi Musa, the Bidūl storyline distances the Layāthnā from Petra geographically. As Figure 4 reveals, āl-Hay is also close to ‘Ain Musa, the spring that served as the chief water source for the indigenous tribes of the entire Petra region before the spread of tourism. This detail, along with Participant 7’s assertion that the Bidūl “were the only ones” in Petra and relied on “the grace of God” to increase their numbers, strengthens the group's connection to the city by emphasizing the willingness of tribal ancestors to endure harder living conditions to reside in Petra. By spatially distinguishing their Petra-dwelling ancestors from the far-off Layāthnā, the Bidūl leverage their well-documented history of living in Petra’s caves and their continued working presence in the city to support their claim that their presence in the ancient Nabataean capital precedes that of the Layāthnā.
Figure 4: This map displays the locations of Wadi Musa, Petra, Umm Sayhoun, ‘Ain Musa, and al-Hay. It also depicts Layāthnā commercial areas (yellow) and Bidul commercial areas (red), as well as The Back Way (blue).

Due to Wadi Musa’s distance from Petra, oral histories propagated by members of the Layāthnā position their ancestors as heroes who protected Petra from Ottoman influence before the Hashemites came to power in Jordan. Participant 55, a Layāthnā sheikh, explained,

The Turks did not enter Wadi Musa. They only made it to that mountain over there [he pointed to the mountain that separates Wadi Musa from the Shobak Plateau]. The tribes gathered and, by the grace of God Almighty, the Turks did not make it any further. A large number of them died, and they did not enter the region. This occurred during the final years of the Ottoman Empire. This period was characterized by tyranny and oppression, so the people revolted; the sheikhs of Wadi Musa refused these politics, which led to the bloody battle that occurred. (Participant 55).

Participant 49, a Layāthnā horse owner, described the importance of this battle to the conflict over tribal access to Petra’s tourism economy today, saying, “A long time ago, Petra was ruled by the Ottomans. The Layāthnā were the ones who fought and removed the Ottomans from Petra. Because of this, Emir Abdallah gave Petra to the Layāthnā. He said, ‘The land of Petra is for you.’” While the stories of Participants 55 and 49 both focus on Layāthnā bravery in fighting Ottoman injustice, the contrast between Participant 55’s characterization that “the sheikhs of Wadi Musa refused these politics” and Participant 49’s assertion that “the Layāthnā were the ones who fought and removed the Ottomans from Petra” reveals the salience of class identity in this storyline. Given the context of decreasing sheikhly influence in Wadi Musa and in Jordan as a whole, the starring role of Participant 55’s sheikhly ancestors in his version of this narrative likely serves the purpose of highlighting Layāthnā chiefs importance in leading their tribe to victory, and with it, cementing the group’s claim to Petra’s tourism-related financial
benefits today. Another noteworthy aspect of the positioning statements of Participants 55 and 49 is the absence of the Bidūl tribe from Petra’s pre-Hashemite history. In contrast, Bidūl narratives about this time period position the Layāthnā as having lived “in āl-Hay,” a detail that reveals that the tribe defines itself in this period in relation to the Layāthnā. The need for this dialogic “other” in Bidūl positioning statements becomes even clearer in the historical narrative of Participant 23, a Bidūl sheikh:

A long time ago, the Wadi Musans were with Auda Abu Taya, who was working with Sharif Hussein. Auda was the paramount sheikh of the Huwaytat. Everything from Arar [Saudi Arabia] to the end of Wadi Araba belonged to Auda Abu Taya, and he commanded 350,000 men...The rights to Petra return to Auda Abu Taya, and only him. (Participant 23).

In asserting that “Wadi Musans were with Auda Abu Taya,” who truly possessed the “rights to Petra,” Participant 23 questioned the Layāthnā tribe’s historical sovereignty over the city, as well as the group’s claimed instrumentality in protecting it from Ottoman oppression. In the process, however, he also reaffirmed the centrality of this delegitimization to his tribe’s historical narratives, revealing them to be subaltern. By refusing to engage the discourses of the Bidūl, on the other hand, the Layāthnā’s dominant narrative establishes their hegemony over a tribal past from which this subaltern group has been eliminated entirely. As Shryock (1997) argued in revealing the dominance of ‘Adwani oral histories over those propagated by the ‘Abbadi tribe, the Layāthnā’s failure to acknowledge the Bidūl’s role in Petra’s pre-1921 history demonstrates a refusal to accept the equality of honor implied by the Layāthnā’s centrality to the narratives of the Bidūl.

The Bidūl only feature in Layāthnā narratives pertaining to Petra’s history following the establishment of Hashemite rule over Jordan. Specifically, the Layāthnā attempt to delegitimize the Bidūl by positioning them as nawwar [gypsies] and shahādīn [beggars] who do not even display karam [generosity] and dayāfa [hospitality], the most basic customs of the Bedouin heritage they supposedly represent. Participant 49, a Layāthnā horse owner, revealed,

The Bidul came to Petra between 1935 and 1940. Before that, they lived in al-Hamayma. The Bidul are not real Bedouins because they ask for money from tourists and do not live in camel-hair tents. The Bedouins are known for their karam and dayafa. Living in a cave does not make you a Bedouin, especially if your livelihood is based off of tourism. (Participant 49).

In addition to the claim that the Bidūl do not practice these basic traits, several participants highlighted that the Layāthnā exhibited karam and dayāfa to the Bidūl when they first came to Petra. As Participant 42 commented, “when the Bidul came, we treated them like neighbors, allowing them to live alongside us in the caves of Petra.” By spreading a counternarrative to the Bidūl’s assertion that they are the original inhabitants of Petra, the Layāthnā simultaneously cast doubt on the Bidūl when they first came to Petra. As Participant 42 commented, “when the Bidul came, we treated them like neighbors, allowing them to live alongside us in the caves of Petra.” By spreading a counternarrative to the Bidūl’s assertion that they are the original inhabitants of Petra, the Layāthnā simultaneously cast doubt on the Bidūl’s origin claim and the tribe’s UNESCO-recognized Bedouin credentials, positioning them instead as a group of poor settlers looking to strike it rich through a stable, sedentary lifestyle.

In the eyes of the Layāthnā, their claim that
the Bidūl arrived in Petra only recently demonstrates why even the current situation of uneven development favoring the Layāthnā is unjust. Given the Bidūl’s alleged status as guests on Layāthnā land, the Layāthnā believe that illegal Bidūl souvenir shops inside Petra in particular give the Bidūl an advantage in the tourism trade that they do not deserve. Participant 53, a qānūnī souvenir seller, encapsulated the frustration that many Layāthnā feel towards the Bidūl’s economic activity within Petra by stating that “those people are governed by a different God; not one of them pays a red cent. There are no more than five permitted shops in Petra. There are 200 non-permitted shops.” Although the glut of souvenir sellers in Petra makes these Bidūl ventures unprofitable, their mere presence is viewed by the Layāthnā as a violation.

The Bidūl, meanwhile, refute the Layāthnā claim that they are lawless migrants by emphasizing that they play an indispensable, albeit informal, role in protecting Petra and the city’s visitors. Participant 5 argued, “the Bidūl have, how can I say, a sense of duty, like when somebody gets lost, the Bidūl go and look for him.”Participant 7 gave a specific example: “Twenty years ago, there was a foreigner who climbed Jabal Umm al-Biyāra, fell, and died. Nobody could get him. The Bidūl got him. The PDTRA and the police could not even get him.” This story rebuts the Layāthnā notion that the Bidūl do not exhibit karam. Moreover, the Bidūl's positioning of the well-resourced and well-funded authorities as unable to perform a basic service, leaving the Bidūl to provide it out of a “sense of duty,” also carries a critique of the government institutions that now lead Petra’s development, implying that the tribe could perform the work they do and more if its members were just left to their own devices.

**Outside Intervention in Petra’s Development**

In contrast to study participants’ accounts of Petra’s tribal history, which feature primarily tribe-based justifications for who should benefit from tourism, class constitutes the primary distinguishing factor between study participants’ interpretations of the role of local and national government institutions in Petra. These class-differentiated Bidūl and Layāthnā positionings of the PDTRA, the state, and their policy initiatives in Petra reveal groups within the local moral order that are in danger of being forced into frozen positions.

Before exploring Bidūl and Layāthnā reactions to specific government interventions in Petra, it is important to examine how members of both tribes view the political institutions that are primarily responsible for creating and carrying out tourism development policy in Petra: the PDTRA and the national government. As the literature review revealed, tribespeople’s opinions of the PDTRA and the national government are influenced by pre-positioning acts consisting of past interactions with these institutions. These experiences, in turn, inform Bidūl and Layāthnā positioning acts towards these authorities’ policies in Petra today.

As discussed in the results section, entrepreneurial-class Bidūl and Layāthnā believe that wāsita is central to finding work at the PDTRA. Many participants also claimed that the PDTRA marginalizes local voices in the creation of development policy. Participant 13, a mukhālaf Bidūl souvenir seller, noted that “the government comes in and destroys the shops that are non-permitted. Clearly, they do not want the shops in Petra. The government should come and talk to people and see what they want.” Participant 41, a Layāthnā hotel manager, added, “The PDTRA is terrible. What happens is that the moment someone gets an important job, he stops caring about his community. I have good ideas about how to improve tourism development while simultaneously helping the local community, but there is no place for me to present these ideas.” Participants 13 and 41 both desire that the PDTRA to “talk to people,” allowing locals to have a voice. Instead, they see a local government that unilaterally “destroys the shops” and has stopped caring about the community.

Many Bidūl and Layāthnā believe that the national government in Amman enriches itself at the expense of Petra’s tribes. Participant 29 asserted, “The national government is only interested in the money that comes from Petra.” Participant 44, a tour guide, concurred, stating that “the ticket price of 50 dinars gets divided up as follows: eight dinars for the horse owners, two dinars for the tour guides, and one for the cleaning staff. The rest goes to the traitors in Amman.” Participant 21, a mukhālaf souvenir seller, elaborated, “The local community should be the one that benefits from Petra financially. It should not be the state. The state has taken the economic benefits and has not returned anything to the local community.” Participant 21’s assertion that the local community, not the state,
should be the primary beneficiary of tourism to Petra illustrates how the Bidūl and Layāthnā are naturally skeptical of the government’s motives in Petra. The thieving and traitorous government behavior alleged by Participants 29 and 44 only compounds these suspicions.

In contrast to the sharp entrepreneurial-class criticisms of the state’s interests in Petra, Layāthnā sheikhs offered more measured critiques. Participant 51 said, “the Ministry of Tourism is the doctor for the chaos that is currently going on in Petra. The main problem is that the government and the Ministry of Tourism do not see the situation as it is; they sit in Amman and make decisions far away from the local community.” Although Participant 51 prefaced his critique of the government’s distance from the local community by exalting the Ministry of Tourism as the “doctor for the chaos” in Petra, it is clear that across class boundaries, local community members expect very little from the national government by way of meaningful efforts to improve their lives.

In addition to complaints about government ineffectiveness, entrepreneurial-class Bidūl and Layāthnā positioned both local and national institutions as highly corrupt. Participant 21 asserted, “The central government does not do anything. The parliament is corrupt.” Participant 43, a Layāthnā hotel owner, concurred, “The problem, and the reason why our situation never improves, is the issue of corruption in Jordan. Every single citizen is corrupt and would become corrupt if given the power. For that reason, nothing ever changes: no ideas get implemented. If there is a project to build a street that should cost 100,000 dinars, they will say that it costs 400,000 and pocket the difference. This is true of all of Jordan, and is something that extends from the PDTRA to the government in Amman, to all levels of society (Participant 43).

The example of the “street that should cost 100,000 dinars” was also mentioned by seven other Layāthnā entrepreneurs, indicating that members of this class invoke this metaphor to illustrate the extent to which corruption pervades their society.

The belief that “every single citizen is corrupt” and that power inevitably breeds corruption reveals the extent to which members of both tribes distrust their governing institutions. Moreover, allegations of government facilitation of corruption only further this narrative. In May 2016, tens of Wadi Musans rioted to demand government action in response to revelations that Azmy Nasarat, a Layāthnā tribesman, and his cronies had conned Wadi Musans out of millions of Jordanian Dinars (JD) through fraudulent real estate investments.58 Participant 52, a qānūnī souvenir seller, claimed that “the state was in on all of this. They saw that it was wrong, but they did not interfere.” Participant 29 added, “Of the seven people implicated in the case, only one went to prison.”

In August 2017, the Jordan Times revealed that a restaurant near the Petra Visitor's Center in Wadi Musa had charged a customer sixty Jordanian Dinars (JD), approximately eighty U.S. Dollars, for a watermelon. The restaurant’s monthly operating costs, however, stood at JD 14,000, leaving no choice but to sell food at inflated prices.59 According to the Times, the restaurant’s landlord is the PDTRA, “which leased at minimal price (JD 600 a year) to the Capital of the Nabataeans Cooperative Society (CNCS), which, in turn, leased the facility at JD 78,000 per year to [the] owner.”60 It is easy to imagine that CNCS’s decision to exact a 130-fold increase on the rent of a government-owned property could fuel local distrust of institutions. However, the magnitude of this breach of trust is exacerbated because, according to Participants 45, 46, and 47, who work at CNCS, the organization’s “primary goal is to help the local community” and “government support was given to Capital of the Nabataeans to get the organization off its feet.” These incidents of local-level corruption lend credence to misgivings entrepreneurial-class Bidūl and Layāthnā, in particular, display toward government institutions, a reality that authorities must confront as they implement their development agendas in Petra.

Class identity and beliefs among Petra’s local communities concerning institutional ineffectiveness, disinterest, and corruption inform Bidūl and Layāthnā reactions to local development policies. This is the case with the Back-Way Project, a recent PDTRA effort to build a paved road from Umm Sayhoun to Qasr Bint al-Fara‘ūn monument in central Petra to allow tourists to access the city by bus, eliminating the need to hike the kilometer-long Siq to arrive at Petra’s monuments. While the road is fully paved and ready for use, it has not yet been used for its original purpose amidst wide-
spread local opposition to the project. Positioning acts from across Petra’s socioeconomic spectrum reveal the class-specific reasons tribespeople support or oppose the project and why PDTRA efforts to complete the project without local input created a frozen position for entrepreneurial-class Bidūl.

According to Wadi Musa’s mukhtār and one Layāṭnā sheikh, the Back-Way project will help the local community. He asserted,

The Back Way will give Wadi Musans money. It will especially help people like the taxi drivers, who will earn a few extra cents driving people from Qasr Bint al-Fara’un to Wadi Musa. Members of the Layathna will also work as bus drivers. The people who are complaining are the Bidul. They see the Back Way as a threat to their work on the donkeys (Participant 42).

Participant 55 confirmed, “The bus companies that will drive on the Back Way will employ people from the local community.” The mukhtār in particular positioned local opposition to the Back Way as an intertribal problem. In his estimation, the Bidūl “see the Back Way as a threat” because they foresee that it will impact their livelihoods. In particular, the Bidūl are wary that this project will reduce demand for donkey rides because tourists will avoid the long walk through the Siq to Petra.

Despite the mukhtār’s confidence that the Back Way would bolster the financial fortunes of Wadi Musan bus and taxi drivers, entrepreneurial-class Layāṭnā widely opposed the project, positioning it as a threat to their tourism-related businesses. Participant 52, a legal souvenir seller, argued, “The Back Way does not benefit the local community. People will only come for one day. They will not stay at hotels or eat at restaurants in Wadi Musa. The person who benefits from this project lives in Amman; he is the person that cuts a deal with the bus company.” Participant 48, a hotel owner, added, “The Back Way was designed without our input. If the PDTRA allows a tourist bus to enter Petra via the Back Way, we will create a problem.”

Beyond threatening the incomes of entrepreneurial-class Bidūl, the Back Way also symbolizes the marginalization of the tribe as a whole in the development process. This reality is demonstrated by Participant 23, a sheikh, who stated, “The government built the Back Way and finished it, but the Bidūl are against it. If they do bring buses to Petra, there will be a battle involving the Bedouins and there will be violence.” In addition to signaling his shared opposition to the Back Way, Participant 23 advocated for resolving this issue through the same means advocated by entrepreneurial-class members of his tribe: violence. In the same way that entrepreneurial-class Bidūl’s claims that they will resort to violence betray their lack of input in government decision-making surrounding Petra’s development, the sheikh’s repetition of this threat reveals that he too occupies this frozen position, confirming Shoup’s (1990) observation about the Bidūl sheikhs’ lack of authority. In August 2019, the loss of agency portended by this breakdown of the wāsīta-based means of representation available to Bidūl entrepreneurs led a Bidūl tribesman to shoot at an empty PDTRA bus as it travelled through Umm Sayhoun towards the Back Way. Particularly against the
backdrop of Bidūl and Layāthnā storylines of outside intervention that indicate local marginalization along class-based lines, this violent incident demonstrates the importance of taking into account the perspectives of all segments of Petra's community. Within the system of local representation currently in place in Petra, this can only be achieved by ensuring that mukhtārs and sheikhs possess the power and sense of responsibility towards their tribal constituencies to effectively fulfill their duties as intercessors and communicators of Bidūl and Layāthnā interests.

Conclusion

At the beginning of this study, I set out to understand the prevalence of tribal identity in Bidūl and Layāthnā discourses around conflict over the distribution of tourism revenues in Petra. Through positioning theory analysis, I identified tribal narratives and discourses related to outside intervention by local and national authorities that are salient across both class and tribal boundaries in Petra's indigenous community. I found that tribal identity plays a significant role in informing interviewees' positioning acts related to tribal narratives, which describe historically-grounded beliefs about how these development resources should be distributed. In contrast, I discovered that class affiliation supersedes tribal identity as a differentiating factor between positioning acts related to outside intervention in Petra's development by local and national government bodies. I also observed that in Petra's entrepreneurial class, interpretations of these policies are heavily influenced by perceptions that these authorities are ineffective, corrupt, and uninterested in prioritizing local community development.

The extent of the salience of class identity in Bidūl and Layāthnā interpretations of state-led tourism development policies in Petra represents a novel finding of this study. During his 1980s fieldwork in Petra, Shoup noted that the increasing viability of tourism as a source of income for the Bidūl and the Layāthnā influenced members of both tribes to engage in various forms of tourism-related entrepreneurship, including knitting handbags as souvenirs, renting horses and camels to visitors, and opening hotels. At the same time, however, he observed that despite the emergence of the tourism-related occupations that constitute the entrepreneurial class today, tribal identity represented the key differentiating factor between development outcomes associated with government interventions in Petra's local community. By expanding the focus of Shoup's (1985) study to include local tribespeople's interpretations of tourism development policy, my positioning analysis revealed the importance of class identity in the creation of both tribal histories and storylines related to outside intervention in the Petra region. For policy-makers concerned with improving access to Petra's tourism economy for members of the city's tribal communities, class-based differences in positioning acts towards initiatives like the Back-Way project reveal the deficiencies associated with perceiving the Bidūl and Layāthnā as tribal monoliths.

Beyond Petra, further fieldwork in Jordanian tourist destinations like the Eastern Desert or Wadi Rum would reveal whether class constitutes a key distinguishing factor between local interpretations of tourism development policies in these contexts as well. While Petra, being a tribal area, is a critical case in this regard, observing the salience of class in the Eastern Desert or Wadi Rum, two tribal areas that have experienced much less tourism development than Petra, would demonstrate the degree to which state involvement in a tourist destination's management facilitates the emergence of class-based inequalities. By identifying at which stage in the development process inequalities begin to appear, the Jordanian government could adjust its tourism sector growth policy to ensure that outside intervention involves local communities instead of distorting them.

On the policymaking and policy implementation level, my findings demonstrate that achieving cross-class indigenous participation in the development process necessitates a shift in how government actors view the sociodemographic makeup of a tourist destination. It is not enough to simply recognize the existence of “six [tribal] communities in Petra,” as Participant 56, an advisor to the Deputy Chairman of the PDTRA's Board of Directors, observed. Rather, bureaucrats at the local and national level must acknowledge that these “six communities” are divided along socioeconomic lines into multiple subcommunities. Ignoring this reality makes the inequality that permeates Petra's tribal communities inevitable and accelerates the development of frozen positions. In the case of the Bidūl entrepreneurial class, the August attack on the PDTRA bus signifies that these losses of agency, which
prevent actors from negotiating their positions using means accepted by the local moral order, have already developed.

To create a development environment where actors enjoy the agency to engage in the positioning process without having to resort to violence, Jordanian policymakers must move away from the monolithic notion of “tribe.” To accomplish this in Petra, PDTRA leadership must incorporate the reality of class-differentiated interpretations of its policy initiatives into its framework for local community representation. Currently, Bidūl and Layāthnā entrepreneurs have to rely on wāsita-based intercessors, like tribal sheikhs and mukhtārs, to protect their interests in Petra’s tourism economy. This study has shown, however, that the agendas of these representatives often contradict those of their tribal constituents, creating gaps in local involvement in the tourism development process. The PDTRA can rectify this shortcoming by establishing alternative, class-based frameworks for tribal representation that complement these traditional channels. For instance, the PDTRA could add seats to its Consultancy Council, an assembly of Petra’s tribal and governmental leaders that convenes to discuss local opinions of tourism development policies and proposals. Specifically, the council could be expanded to include Bidūl and Layāthnā representatives for souvenir sellers, horse and donkey owners, refreshment stand owners, and other occupations within the entrepreneurial class. It is only through involving the class-based subgroups that exist within Petra’s “six tribal communities” that government authorities can achieve a coherent tourism development policy that features widespread local participation in the city’s growth.

Endnotes

2. Participant 23 (Bidūl tribesperson) in discussion with the author, March 2019; Participant 29 (Layāthnā tribesperson) in discussion with the author, March 2019.
13. Ibid., pp. 16–17.
18. Ibid.
19. Ibid.
20. Andrew Shryock, Nationalism and the Genealog-

21. Ibid.

22. Ibid.


24. Ibid., p. 211.


28. Ibid.


32. Participant 29 (Layāthnā tribesperson) in discussion with the author, March 2019.


34. Ibid.

35. Ibid.


38. Ibid., pp. 213–214, 218, 246, 258.


43. Ibid., p. 281.

44. Ibid., p. 282.


46. Massad, Colonial Effects, p. 79.


48. Ibid., p. 238.

49. Participant 48 (Layāthnā tribesperson) in discussion with the author, March 2019.


51. Participant 29 (Layāthnā tribesperson) in discussion with the author, March 2019.


54. The George Washington University IRB approved this research under application number NCR191132.


56. Participants 4, 5, 7, 13, 32, 34, 52, 53, and 54 (Bidūl and Layāthnā tribespeople) in discussion with the author, March 2019.

57. Participants 5, 6, 7, 8, 9, 14, 22, 23, 24, 25, 28, 29, 42, 49, and 50 (Bidūl and Layāthnā tribespeople) in discussion with the author, March 2019.


60. Ibid.

Bibliography


Death from Above: Evaluating the Impact of Drone Strikes on AQAP Activity

Chris Conrad, Haverford College

Abstract

How has the U.S. drone campaign in Yemen affected Al-Qaeda in the Arabian Peninsula (AQAP)? This study seeks to answer two questions. First, why might drone strikes increase or decrease AQAP activity? Second, what does the empirical record demonstrate is the case? The study begins by providing background information on AQAP and the U.S. drone campaign in Yemen. I then turn to a literature review of relevant literature on AQAP, the U.S. drone program, and similar counterterrorism programs such as leadership decapitation. The study proceeds to outline a variety of plausible causal explanations for both hypotheses that drone strikes have increased (or decreased) AQAP activity. After presenting causal explanations for both hypotheses, a quantitative analysis is conducted to determine which hypothesis is more convincing. The findings suggest that drones have counterproductive effects.

Introduction

The observation that grounds this study’s question is relatively simple. It seems as if the U.S. drone program has not had a noticeable impact on AQAP—despite President Obama’s increased use of drone strikes in Yemen, AQAP expanded its membership significantly. This is puzzling: if drones were an effective counterterrorism tool and meaningfully degraded the membership and organizational structure of Al-Qaeda and its affiliates, one would expect AQAP to be shrinking or, at the very least, not expanding. Thus, this study’s question is two-fold. First, why might drones positively or negatively impact the ability of terrorist groups to launch attacks? Second, which hypothesis do the data in Yemen support? Is the surface level assessment that drones have not meaningfully affected AQAP correct, or does a more careful look at the empirical data demonstrate the opposite?

This study seeks to add to the literature through its unique focus on AQAP in Yemen. Existing empirical research has focused almost exclusively on Pakistan, and no empirical analysis has been done on the drone war in Yemen. However, recent data indicate that drone strikes have become even more common under the Trump administration, making meaningful empirical analysis possible. Additionally, the ongoing Yemeni civil war has rendered large swaths of the country into ungoverned spaces analogous to the FATA region in Pakistan studied by much of the empirical literature on drone strikes. This study also implicitly tests whether findings in one country can
be extrapolated to another—do drone strikes have the same effect on Al-Qaeda Central (AQC) in Pakistan as they do on AQAP in Yemen, or are the organizations qualitatively different to the degree that strikes have different effects? If strikes have noticeably different effects on AQAP, why might this be the case? Finally, this study adds to the qualitative literature on the U.S. drone war in Yemen by empirically testing their claims.

**Background**

AQAP is an Al-Qaeda affiliate that emerged in 2009 after the merger of two previous affiliates, Al-Qaeda in Yemen (AQY) and Al-Qaeda in Saudi Arabia (AQSA). AQAP has been labeled the most lethal Al-Qaeda affiliate, capable of conducting a domestic insurgency while keeping its eye on the “far enemy.” Both AQY and AQSA emerged after the defeat of the Soviets in Afghanistan and the subsequent fall of the Soviet Union. In 2006, two senior jihadis, Nasir al-Wuhayshi (personal assistant to Bin Laden) and Qasim al-Raymi escaped from prison and became top leaders in AQY. Al-Wuhayshi announced the merger between AQSA and AQY in 2009 and became the emir of the organization, with al-Raymi serving as AQAP’s chief military commander. By this time, the Saudi government had cracked down on AQSA, arresting or killing its leadership, effectively pushing AQSA underground. Thus, the merger made strategic sense. After al-Wuhayshi was killed in 2015, al-Raymi became the emir of AQAP. AQAP’s funding comes from a variety of sources, including the drug trade, illegal taxes, bank robberies, hostage ransoms, and donations from private individuals and charities in Saudi Arabia whose fundamentalism aligns with AQAP’s goals. Prior to the outbreak of the Yemeni civil war in 2015, AQAP’s attacks largely targeted the Yemeni government (under both Ali Abdullah Saleh and Abd Rabbuh Mansur Hadi). AQAP also occasionally attacked against western nations, notably the 2000 bombing of the U.S.S. Cole (executed by AQY), the failed 2009 Christmas Day bombing—where an AQAP operative attempted to detonate a plane flying to Detroit, Michigan—and the Charlie Hebdo attack in France. AQAP’s stated goals are in line with Al-Qaeda’s broad goals of purging Western influence from Muslim countries and overthrowing secular apostate governments in the region to replace them with Islamic regimes. AQAP’s specific strategic goals include overthrowing the government in Sana’a, assassinating members of the Saudi royal family and western nationals in the region, and launching attacks against the U.S. and its allies. However, as will later be discussed, AQAP is also strongly opposed to Shiite Houthi rebels which primarily reside in the northern region of Yemen, and has at times had a complex relationship with the government in the south.

Of note is AQAP’s strong propaganda apparatus—it regularly publishes an online English magazine titled Inspire, aiming to create homegrown radicals in the U.S. to carry out attacks. This is fairly unique among jihadi groups, the Islamic State being the other notable example of a group publishing an English magazine aimed at creating homegrown attackers. Despite the killing of Anwar al-Awlaki and Samir Khan (AQAP’s propaganda leadership responsible for the magazine) in a 2011 U.S. drone strike, Inspire is still published, though at a significantly decreased rate. Domestically, AQAP’s propaganda has tailored the jihadi narrative to the Yemeni context, and, taking into account the high illiteracy rate, is often elaborated through oral culture and traditional customs. Additionally, AQAP publishes an Arabic magazine tailored to the Yemeni audience online.

Despite its partial origins in Saudi Arabia, AQAP is almost entirely located in Yemen. A confluence of factors explains this decision. The Saleh government in Yemen (prior to Saleh being replaced by his vice president Hadi during the Arab Spring) chose not to take significant actions against AQAP because the organization’s presence could be used as a bargaining chip to secure U.S. assistance. Furthermore, Saleh had a history of permissive alliances with Sunni militant groups to defeat communist and Shiite domestic opponents, particularly the Houthi movement. Under Hadi’s government, Yemen presented a comparatively preferable environment for AQAP compared to the high capacity, repressive Saudi state as a result of rampant corruption in Hadi’s administration and a weaker state security apparatus. Since the Houthi takeover of the capital (resulting in Hadi’s subsequent flight to Saudi Arabia) and the escalation of the Yemeni civil war in 2015, AQAP has faced a comparatively more permissive security environment due to its primary opponents (the government and the Houthis) being focused on fighting each other, allowing it to rapidly expand, as discussed later.
The use of drone strikes by the U.S. began under the Bush administration as part of the global war on terror (GWOT) after 9/11, but escalated significantly under the Obama administration. The first drone strike was launched in 2002 by the Bush administration in Yemen. Notably, drones have been used both in conventional warzones (Afghanistan, Iraq, Libya) and unconventional warzones in which U.S. troops are not directly involved (Yemen, Somalia, Pakistan). Under the Trump administration, drone warfare has escalated. President Trump has reportedly removed checks established by President Obama to limit drone strikes, and is increasing the number of drone strikes compared to President Obama, particularly in Yemen. For example, in early March of 2017, President Trump conducted an unprecedented forty drone strikes in the span of a few days in Yemen.

Most justifications for the expansion of the U.S. drone campaign have a few common themes. The central argument is that drone strikes are effective at constraining terrorist’s ability to operate (and conduct attacks against the U.S. homeland). As Daniel Byman argues, “by killing key leaders and denying terrorists sanctuaries in Pakistan, Yemen, and, to a lesser degree, Somalia, drones have devastated al Qaeda and associated anti-American militant groups.” Put simply, drones are thought to “trim the grass” by killing terrorist operatives. However, criticism of the program is common, grounded on legal and ethical questions about a program that may constitute extrajudicial killing with little transparency or oversight abound. Moreover, critics point to civilian casualties caused by drones both as a moral argument against the program and as an argument that the program is ineffective, insofar as civilian casualties may contribute to boosting the recruiting efforts of terrorist groups.

Why use drones instead of alternatives? While Special Operations forces have conducted raids in permissive political contexts such as Iraq (pre-2014) and Afghanistan, other environments (particularly in Pakistan, Yemen, and Somalia) present challenges to the use of raids, making drones a useful tool. Additionally, the post-Iraq War unpopularity of devoting “boots on the ground” to fight terrorism make drones a politically salient option to continue the GWOT. Finally, drones are relatively inexpensive and do not pose any risk to the lives of U.S. troops.

The public controversy on the use of drones often centers on civilian casualties. Critics argue that drones result in unacceptably high levels of civilian casualties as collateral damage. In addition to moral concerns surrounding civilian casualties, critics suggest that civilian casualties generate resentment in target populations, helping terrorist groups recruit more to their cause. Others disagree, arguing figures about civilian casualties are inaccurately biased and that polling data about what target populations think of drones is unreliable and unrepresentative. Additionally, some argue that in addition to the severe problems with finding reliable data about civilian casualties, drones are comparatively more discriminating in their targeting than other options such as air power, ground intervention, or asking local governments with poor human rights records to hunt terrorists instead.

Other critics emphasize the impacts of drones on local civilian populations that go beyond direct casualties. One detailed report conducted numerous interviews with Pakistani civilians to document the negative impact of drones on people's economic, psychological, educational, and other forms of well-being. However, Christine Fair argues that the aforementioned study on Pakistani civilian mental health is flawed due to sampling bias, methodology problems, failure to capture the full spectrum of views on the ground, failure to disambiguate the causes of psychological harm, and inaccurate attribution of killings by the Pakistani military to drones.

While the public debate on the impact of drones on civilian populations is important, this paper primarily focuses on examining the available data to determine if drones are even an effective counterterrorism tool. This is because critical questions about the effectiveness of drone strikes remain unanswered. Unfortunately, reliable, systematic data about terrorist group recruitment does not exist, which makes directly testing recruitment arguments in a systematic manner impossible—at best, the evidence is anecdotal. Systematic empirical research on the U.S. drone campaign has been relatively limited, with a few exceptions. Because drones will likely continue to serve as a counterterrorism tool for years to come, analyzing the effects of drone strikes on terrorist activity (particularly the frequency and lethality of attacks) is critical.
Literature Review

Four literature bases are relevant to this study’s topic. First, the empirical literature on leadership decapitation is of some relevance to questions of drone effectiveness, as drones ostensibly target terrorist leadership figures. Second, existing empirical literature about U.S. drone strikes outside of Yemen provides plausible causal explanations for both hypotheses tested by this study, as well as useful information for study design. Third, limited qualitative literature documenting the effect of drone strikes on AQAP provides additional explanations for the causal mechanisms justifying each hypothesis. Fourth, recent literature about AQAP—particularly how the group was affected by the 2011-12 Arab Spring and the civil war in 2015—provides a qualitative picture of the current conflict that helps inform variable selection for this study’s empirical analysis.

Empirical literature on decapitation

Existing empirical literature on decapitation is of mixed relevance to this study’s topic. Much of the justification for the U.S. drone program has been based on the argument that drone strikes “decapitate” terrorist groups. That is, eliminating key leadership figures hastens a group’s demise. However, research has found mixed results on the effectiveness of decapitation. Prior to two studies conducted in 2012, the consensus among scholars was that decapitation was generally ineffective. Lisa Langdon’s early study found decapitation was ineffective but used too small a sample size to be statistically significant; Aaron Mannes’ study had mixed results on a relatively small, and largely did not find statistically significant results. Jenna Jordan’s 2009 study was the first to conduct a large-N study of decapitation strikes, and found decapitation to be ineffective. Specifically, Jordan found that decapitation was only effective in 17 percent of cases, and that typically, decapitated organizations declined at a slower rate than terrorist organizations that did not experience decapitation. Additionally, decapitation was less effective in “older, larger, religious, and separatist organizations.”

However, some recent literature has diverged. Bryan Price’s 2012 study of leadership decapitation found that leadership decapitation had a significant effect on group mortality—decapitated groups were 3.6 to 6.7 times more likely to end than non-decapitated groups. Decapitation achieved greater results when it occurred earlier in a terrorist group’s life cycle. Interestingly, Price found that all methods of decapitation (killing, capturing, or capturing and then killing) resulted in higher group mortality rates, and that normal leadership turnover absent decapitation had similar effects. Contrary to Jordan, Price found that larger organizations were not more resilient, and that religious terrorist groups were less resilient to decapitation. Patrick Johnston also found decapitation to be effective in his 2012 study. Specifically, Johnston found that leadership decapitation “(1) increases the chances of war termination; (2) increases the probability of government victory; (3) reduces the intensity of militant violence; and (4) reduces the frequency of insurgent attacks.” Additionally, as Manna points out, numerous studies have examined the Israeli targeted killing policy to test if decapitation is effective, with mixed findings.

Max Abrahms and Jochen Mierau conducted an empirical study of decapitation’s effects on terrorist tactics, examining the West Bank/Gaza Strip and Pakistan/Afghanistan. They found that successful decapitation resulted in militant groups turning to target civilians instead of military targets. They argue that their findings suggest that this result was a product of lower-level members who had ascended to leadership positions having less restraint when targeting civilians compared to previous leadership. This raises an interesting point, which is that even if decapitation may successfully bring about the demise of terrorist groups, it may induce short term increases in attacks against civilians. In fact, Abrahms and Phillip Potter found in another empirical study of decapitation that drone-induced decapitation in Afghanistan and Pakistan caused increased targeting of civilians, though only when the drone strikes successfully decapitated a leader.

However, decapitation literature may not capture the realities of the U.S. drone program, particularly in Yemen. As Manna points out, Al-Qaeda has gradually become more decentralized over time, making decapitation unlikely to result in group mortality. AQAP, in particular, is becoming more decentralized over time, partially in response to counterterrorism pressure after the death of al-Wuhayshi in a 2015 drone strike. Additionally, if Jordan’s findings are to
be trusted, AQAP’s relative age and religious nature could be expected to make it resilient to decapitation. More importantly, the U.S. drone program is not solely focused on decapitation. In Yemen, the U.S. has turned to using signature strikes, in which individuals or groups are targeted not based on specific intelligence (targeted killing of a particular leadership figure), but rather on whether they fit the “signature” of a militant.\(^{60}\) As a result, drone strikes have targeted not only AQAP leadership, but rank and file members, making the decapitation literature questionably relevant.\(^{61}\)

One estimate found that U.S. drone strikes in Yemen killed one high-value individual (HVI) for every sixty deaths.\(^{62}\)

**Empirical literature on drone strikes outside of Yemen**

Limited empirical research exists to examine the specific impacts of drone strikes.\(^{63}\) Of the research that has been done, findings have been mixed. Patrick Johnston and Anoop Sarbahi’s study of drone strikes in the FATA region of Pakistan found that strikes reduced the incidence and lethality of terrorist attacks in the week following a strike.\(^{64}\) Additionally, they found that the killing of HVIs reduced terrorist attacks, and that terrorist activity did not shift into proximate regions as a result of drone strikes.\(^{65}\) However, the violence reducing effect of drone strikes largely disappeared in subsequent weeks.\(^{66}\)

David Jaeger and Zahra Siddique conducted an empirical study of drone strikes’ impact on Taliban activity in Pakistan and Afghanistan, obtaining mixed results. In Afghanistan, drone strikes had no effect on Taliban activity.\(^{67}\) However, in Pakistan, there was a vengeance effect after the first week, followed by a deterrent/incapacitation effect in the second week.\(^{68}\) The effect also varied by the region in which a strike took place, as different factions of the Taliban have control in different regions.\(^{69}\)

Megan Smith and James Igoe Walsh examined the empirical effect of drone strikes on Al-Qaeda’s propaganda output. They found that drone strikes did not have a significant effect on propaganda output immediately following a strike, and any observed effects took months to materialize.\(^{70}\) Their findings indicated no strong overall relationship between drone strikes and propaganda output.\(^{71}\)

Emily Manna examined the effect of drone strikes in Pakistan on terrorist attacks.\(^{72}\) Manna’s model examined the longer term effects of drone strikes using a difference-in-difference model.\(^{73}\) Manna’s findings indicated that drone strikes increased terrorist attacks, though the statistical significance of drone strikes diminished when the model used a lagged dependent variable.\(^{74}\)

More recently, Asfandyar Mir and Dylan Moore conducted an empirical study of the effect of the U.S. drone program in Pakistan’s FATA region. Notably, Mir and Moore’s analysis examined the holistic effect of the drone program—as opposed to the effects of individual strikes, which is what most studies to date have examined—and found that the violence reducing effect was greater than the sum of its parts. Indeed, aggregating the effects of each individual strike only accounted for 25 percent of the reduced violence.\(^{75}\)

Specifically, they isolated anticipatory effects as key to the reduction in violence: insurgent groups, knowing that the drone program was in place, anticipated that there was increased risk to their security, and adjusted their behavior in ways that limited their ability to operate. For example, groups avoided use of communications devices, avoided predictable movement patterns that would result in a drone strike, and more.\(^{76}\)

Other empirical research has examined public opinion of drone strikes in target countries, which is relevant to the common argument that drones boost terrorist group recruitment and thus the frequency of their attacks. Karl Kalthenthaler, William Miller, and Christine Fair examined a Pew Research survey of Pakistanis that demonstrated overall unfavorable opinions of U.S. drone strikes.\(^{77}\) Their study sought to explain what characteristics made people more or less likely to support drone strikes. Unsurprisingly, Pakistanis who viewed themselves as religious fundamentalists and saw the U.S. as the enemy were most likely to oppose drone strikes. On the other hand, those who thought that the U.S. favors Pakistan over India were more likely to support strikes.\(^{78}\) Additionally, higher levels of education correlated with greater support for drone strikes.\(^{79}\)

James Igoe Walsh empirically examined how people responded to the possibility of civilian casualties caused by precision weapons, such as drones.\(^{80}\) Walsh examined whether the supposed precision of the weapons would raise people’s expectations, thereby making civilian casualties a more salient concern.\(^{81}\)
Walsh’s findings demonstrated that precision weapons, particularly when weapons were unmanned (drones), initially increased support, but that civilian casualties and civilian harm reduced support for precision weapons.82 However, Walsh’s study was of the U.S. public, and is thus of limited relevance to this study’s research.

Perhaps most importantly, Aqil Shah recently attempted to directly study whether drone strikes in Pakistan had produced blowback.83 Based on original interviews, Shah found no evidence of blowback in the local areas in which drone strikes took place.84 Interviewees preferred drones to the alternative, Pakistani military offensives, which produced worse casualties.85 Other interviews demonstrated that radicalization often preceded drone strikes and were due to other factors. Some even said that that blowback claims were largely made for political reasons.86 Similarly, Shah found in interviews with a variety of experts, security officials, and survey data of captured terrorists that drones were not a cause of national blowback (i.e. increased recruitment outside of the territory in which strikes took place), and most attributed radicalization to other political and economic factors.87 However, because Shah’s research was based on field work it potentially suffers from a lack of external validity.

Qualitative literature on drone strikes in Yemen

While all prior quantitative work on drones has centered on Pakistan and Afghanistan, some qualitative research has been done on drones in the Yemeni context. In a relatively short piece, Michael Horton argues that Yemeni civilians despise the drone campaign, and AQAP has used drone strikes as a propaganda tool to boost recruitment.88 Andrew Terrill acknowledges that drones have helped achieve notable successes, including the killing of Anwar al-Awlaki and playing a crucial role in supporting a Yemeni military counter-offensive when Hadi was having severe trouble consolidating control of parts of the armed forces in 2012.89 However, Terrill concurs with Horton, arguing that the Yemeni civilian population strongly dislikes drones because they violate the country’s territorial sovereignty and produce civilian casualties.90 Gregory Johnsen argues that drones have failed in Yemen because unlike in Pakistan, where AQC was an Arab population in a non-Arab country, AQAP is primarily made up of Yemenis in their home country.91 As a result, recruits do not have to travel long distances to receive training and can instead join the cause without having to move. Furthermore, AQAP members killed in strikes are woven into the local social fabric, ensuring people join AQAP to exact revenge on the U.S. for killing their friends and family.92

In contrast to the relatively short pieces outlined above, longer qualitative studies exist too. Christopher Faulkner and David Gray argue that drones have caused resentment among the Yemeni civilian population due to civilian casualties.93 However, they also recognize that drones have killed AQAP operatives and pushed parts of the organization underground, and argue that defeating AQAP requires non-military means as a complement to ongoing drone strikes.94

Leila Hudson, Colin Owens, and David Callen argue that drone strikes produce blowback in the form of retaliatory terrorist attacks, increased recruitment of the civilian population, and destabilization of the Yemeni government.95 In particular, they argue that the use of signature strikes in Yemen has increased recruitment, and, because the Yemeni government publicly supports the drone program, drone strikes likely increase civilian discontent with the government, destabilizing the country.96

Dana Greenfield and Stefanie Hausheer argue that signature strikes in Yemen have produced significantly civilian casualties and that the Obama administration has largely deceived the public about how targeted and precise strikes are.97 They point out that intelligence used to target drone strikes in Yemen is often unreliable for a variety of reasons.98 As a result, even if drone strikes do not cause the civilian population to join AQAP, they make civilians willing to provide a safe haven and generally permissive environment in which AQAP can operate.99

Michael Titus and David Gray argue that drone strikes have been effective at eliminating a number of critical leadership figures in AQAP.100 While their article largely focuses on the unfolding of the Yemeni civil war that had recently begun in 2015, they note that U.S. drone strikes and involvement in Yemen had declined at the time of their writing.101 As a result, they argue that increased drone strikes were necessary to disrupt AQAP’s operations and supply routes, while leadership decapitation could result in infighting to select a new leader.102
Contrary to critics of the drone program, Christopher Swift argues that drone strikes have not produced the expected blowback in Yemen. Based on a dearth of on-the-ground interviews, Swift found that economic factors were the primary motivating factor in explaining why people chose to join AQAP. Additionally, drones did not play a role in recruitment—none of Swift’s interviews indicated drones as the proximate cause for AQAP’s increased membership—as one militia commander put it, “even if drone strikes were to cease … the economic and social conditions that facilitate terrorist recruiting in Yemen would still remain.” However, similar to Hudson, Owens, and Callen, Swift did argue at the time that drones and the public anger they created could undermine the political transition from Saleh to Hadi, as Hadi was actively supporting the program.

Thomas Bolland and Jan Andre Lee Ludvigsen argue that drones have been ineffective in the fight against AQAP. They borrow Javier Jordan’s theoretical framework for assessing the effectiveness of a drone campaign that examines the effect of drones on hierarchical structure, qualified human resources, and key material resources, adding a fourth variable of target correspondence to measure blowback. They argue that, first, AQAP is resilient to decapitation—drone strikes on leadership have been followed by quick and effective succession; second, that AQAP, despite losing human resources, has retained significant capabilities; and third, that drones have not denied AQAP access to funding and weaponry. Additionally, they argue based on anecdotal and qualitative evidence that drones have produced blowback among the civilian population.

Elisabeth Kendall’s recent analysis of AQAP’s organizational health echoes the anticipatory effects argument of Mir and Moore. Kendall argues that increased drone strikes in 2017 have killed a significant number of AQAP members, reducing morale. Additionally, after al-Wuhayshi’s death from a drone strike, al-Raymi has been less charismatic and popular as a leader. More importantly, Kendall isolates how AQAP’s leadership “has gone to ground,” imposing organization-wide bans on mobile and internet communications to avoid detection (presumably by drones), which has made propaganda production and communication more difficult, contributing to fragmentation and infighting. Additionally, drone strikes may decrease rather than increase AQAP’s support from the local population, which may blame AQAP for making them into targets.

Before turning to the literature on AQAP, it is important to point out the lack of consensus within any of the literature outlined above. Among the empirical studies, different models examining the same country produce different results, so it remains critical to test the hypotheses of both sides of the drone debate in a new context to see if any existing findings are generalizable. The qualitative work on drones in Yemen is sorely lacking in empirical grounding and rife with assertions and counter-assertions about blowback based on limited samples (Swift’s interviews) or public opinion surveys that cannot predict whether drones will play a significant cause in recruitment or measure if drones have an actual impact on AQAP’s tendency to conduct attacks within Yemen. As such, an empirical analysis is necessary to test the competing claims in the qualitative literature.

Given that the group’s broad goals and origins are described above in the background, this section primarily seeks to trace AQAP’s recent history during the Arab Spring and Yemeni civil war, with the goal of isolating relevant external factors to use as control variables in the empirical analysis. While it is impossible to perfectly capture every minute detail of the political situation in Yemen that might influence AQAP’s ability or motivation to conduct attacks, the hope is that some useful turning points can be isolated to make the empirical analysis more robust.

Before diving into specifics, a discussion of Christina Hellmich’s article on AQAP is worthwhile. She argues that AQAP has strong incentives to appear “unified, competent, and powerful” even when the reality on the ground may not reflect propaganda—a seemingly stronger AQAP generates greater fear amongst its enemies, making the organization seem more dangerous than it actually is. As a result, gathering reliable data on anything but the organization’s leadership is difficult. Assessing the number of AQAP members is even impossible: the organization’s leadership is difficult. Assessing the number of AQAP members is even impossible: the organization may intentionally misrepresent their numbers to appear stronger, while the Yemeni government has an incentive to exaggerate the threat to secure increased...
military aid. Additionally, there is extreme difficulty in separating active members from sympathizers—the organization cares about secrecy, so “there is no reliable way of identifying who belongs to AQAP.” Thus, it is important to take the qualitative literature on AQAP with a grain of salt. Additionally, the difficulty in distinguishing between members and sympathizers may uniquely increase the risk of blowback to drones, as discussed later.

Given that the vast majority of drone strikes occurred in 2011 and later (only four took place between 2002 and 2010), the first important topic to examine is AQAP’s involvement in the Arab Spring. After Hadi took power from Saleh in 2012 due to widespread protests in 2011, the country descended into a humanitarian crisis—poverty, unemployment, and famine became more widespread. In the midst of the transition, both sides accused each other of affiliation with AQAP. Because the transition from Saleh to Hadi was widely seen as lacking legitimacy and signaled that changes in governance were possible, both the Houthis and southern secessionists managed to make territorial gains in the aftermath, while AQAP struggled to hold territory because it lacked on the ground legitimacy. Despite this, AQAP did manage to capture notable cities and territory in the aftermath of the Arab Spring. AQAP benefited from the transition in other ways—it gave money to tribes to win their support and established Ansar al-Sharia to be their front-facing governance-focused organization to build popular support. Even after a series of government counteroffensives in 2012 and 2013, AQAP was resilient and continued to stage increasingly sophisticated attacks, demonstrating the weakness of the post-Arab Spring state’s security apparatus. However, as noted above, some reports indicate drones played a crucial role in counteroffensives against AQAP’s territorial gains in the period while Hadi was still consolidating power. More importantly, in late 2012, a coalition of tribal militias allied with the Yemeni Army to force AQAP out of territory it had taken in Abyan.

AQAP’s strategy during the post-Arab Spring period was relatively heavy handed, which contributed to its loss of territory in late 2012 and 2013. When AQAP took control of territory in Abyan and Shabwah in 2011, they overturned tribal law and custom, imposing strict interpretations of Islamic law, including a ban on the trade of qat (a narcotic drug) that made up an important portion of the economy. As a result, they lost local support, which played a key role in motivating tribal militias (the critical actor) to kick AQAP out with the assistance of the Yemeni army. As Katherine Zimmerman puts it, AQAP “appears to have learned from strategic errors in 2011 and 2012, which led to a popular uprising against it… al-Wuhayshi, advised his Algerian counterpart in the summer of 2012 that, based on AQAP’s experiences in Abyan, meeting the people’s basic needs was the first step in governance.” Put simply, AQAP experienced a notable change in strategy in late 2012 as a result of their failed governance experiment in the wake of the Arab Spring.

The second major event to examine is the role of the Yemeni civil war in AQAP’s development. The eruption of the Yemeni civil war and toppling of the Yemeni central government by Houthi forces in early 2015 played a significant role in allowing AQAP to expand. Importantly, Saleh appears to have allied with Houthi forces against Hadi, and in the resulting power vacuum from Hadi’s fall, AQAP gained the ability to make significant inroads in Shabwah and Abyan, along with taking nearly complete control of Hadramawt. AQAP was also given ample territory to fill as a result of Saudi airstrikes against Houthi forces that forced them to retreat, allowing AQAP to fill the void in a variety of territories. Additionally, in the midst of war, AQAP was able to increase its numbers by staging a massive jailbreak, finance itself by robbing the central bank, and increase its capabilities by seizing military hardware.

Notably, AQAP laid the groundwork for expansion in late 2014 by using Ansar al-Sharia to reach out to tribes for support, often by purchasing allegiance, as the Sunni-Shia divide is not especially entrenched in Yemen—Sunni tribes do not automatically align with AQAP and oppose the Shia Houthis, they act in their short term interests. This distinction is relevant—support from tribes is often cited as a primary source of AQAP’s resiliency, yet extensive fieldwork in Ma’rib and Al Jawf indicates that AQAP does not have a formal alliance with the tribes. In addition to simply purchasing support from tribes, AQAP has been careful to limit civilian casualties so that they can leverage local resentment against the Houthis to gain support, and has made fighting the Houthis an important organizational goal.
lessly, AQAP left governance to local councils instead of imposing strict Islamic law and conducted some public works projects to amplify goodwill, demonstrating that they had learned from their post-Arab Spring mistakes.133 Particularly in Hadramawt, AQAP’s changed strategy has been notable—they rebranded as the Sons of Hadramawt, did not raise the Al-Qaeda black flag, only intermittently implemented Sharia law, established a council to provide basic services and pay salaries, solved water problems, land disputes, rebuilt infrastructure, and attempted to provide healthcare.134 As Ardemagni points out, part of the motivation for AQAP’s strategic shift towards governance may have been the Islamic State’s attempt to infiltrate Yemen—AQAP took the less hardline approach to prevent defections and get tribal support over the Islamic State.135 It is also worth noting that control of the Hadramawt is relatively important—the governate is the most natural resource-rich region in Yemen and remained relatively less touched by the civil war, causing many IDPs to flee there.136

However, AQAP has declined since its peak territorial control in 2015 during the Yemeni civil war. In 2016, Saudi and UAE forces in tandem with Yemeni government forces began a campaign to push AQAP out of Mukalla.137 While the war is still ongoing, the UAE launched a new counteroffensive in 2018 to recapture territory held by AQAP, which appeared to successfully remove AQAP from large swaths of Hadramawt, Shabwah, and Abyan.138 Despite this, there are reasons to be skeptical—the UAE largely used regional militias that lack a unified chain of command and may be prone to committing human rights abuses, and the militias do not have the manpower or training to consolidate control of large swaths of territory.139 Additionally, AQAP has a history of strategic retreats and is reluctant to fight against superior forces, so it is likely the organization’s capabilities have not been significantly impacted by the UAE counteroffensive, even if their territorial control has.140

Yet, the organization still appears to broadly be in decline. As explained above, Kendall’s report on AQAP demonstrates that the organization is having leadership problems, losing numbers due to counterterrorism efforts, is largely unable to communicate, and is fragmenting as a result.141 In addition, the group is losing popular support from the tribes, as the UAE has been aggressively recruiting to enlist locals in the fight against AQAP.142 The numbers seem to back up the argument that AQAP is in decline—the organization conducted less than half as many operations in the first 6 months of 2018 compared to either 6 month period in 2017, and AQAP’s propaganda output has declined significantly in recent months as well.143

Hypotheses

The empirical section of this study tests two primary hypotheses:

H1: All else equal, drone strikes reduce the frequency and/or lethality of terrorist attacks.

H2: All else equal, drone strikes increase the frequency and/or lethality of terrorist attacks.

A number of causal explanations exist for each hypothesis, largely explained above. Johnston and Sarbahi divide the explanations into two categories: degradation and disruption.144 Degradation suggests that drone strikes will impede a terrorist group’s ability to conduct attacks by decapitating HVIs that are difficult to replace such as leadership figures and those with important technical skills.145 This explanation follows from the empirical literature on decapitation indicating that such strikes are effective at reducing a group’s ability to conduct attacks. Additionally, Mir and Moore outline other “kinetic effects” of drone strikes—in addition to killing HVIs, drone strikes (particularly signature strikes) may result in rank-and-file attrition and the destruction of group capabilities.146 Kendall outlines additional kinetic effects in the Yemeni context—degradation of AQAP’s rank-and-file membership has supposedly decreased morale and increased the pressure for jihadists to defect or flee the organization, while possibly turning civilian populations against AQAP, who are blamed for making civilians into drone targets.147 It may also be the case that drones deter those who join AQAP for economic reasons, knowing that strikes effectively remove whatever material benefits anticipated by joining the group.

Johnston and Sarbahi’s disruption mechanism is similar to the “anticipatory effects” outlined by Mir and Moore. Drone strikes may impede the ability of terrorist groups to organize and operate in a territory by increasing the security risk posed to terrorist
Mir and Moore go in-depth, explaining that the U.S.’s vast intelligence capabilities combined with the perceived threat of drone strikes induce terrorists to change their behavior to avoid communication and patterned movement. They may also cause breakdowns in trust within the group as leaders become paranoid of infiltration. This echoes Byman’s argument that drones impede the ability of terrorist groups to communicate, be in the open, hold large training camps (that could be wiped out in a single strike), and overall turn “command and training structures into a liability, forcing the group to choose between having no leaders and risking dead leaders.” This may well be at play in Yemen—as Kendall notes, AQAP has recently gone largely silent, sending directives to stop mobile and internet communication and significantly reducing propaganda output, ultimately making it difficult for the leadership to exercise control.

Why might drones increase terrorist attacks? First, as Jaeger and Siddique outline, drone strikes may result in vengeance attacks in retaliation, increasing violence. For example, drones were cited as the explicit motivation for attacks in 2009 and 2012 perpetrated by AQAP. Second, decapitation may result in short term increases in violence against civilians, per research outlined above by Abrahms, Mierau, and Potter that leadership figures are replaced by lower-level members with less civilian restraint. Additionally, conflicts within terrorist groups may break out after decapitation as various members fight over who will take the dead leader’s place and result in civilian killing as rivals seek to establish themselves. Third, drone strikes may cause terrorists to attack civilians in an attempt to eliminate individuals who are suspected of having informed on the group (enabling the drone strike), especially when the group is operating in an environment in which state security forces are not immediately nearby to be targeted.

Fourth, and perhaps most importantly, drone strikes may result in blowback—civilians may be radicalized as a result of drone strikes, thereby increasing a terrorist group’s manpower and capability to launch attacks. If drones result in negative impacts on the civilian population, civilians may become more sympathetic to terrorist groups and join them to exact revenge. Additionally, drone strikes may enable terrorist groups to produce propaganda about the strikes, painting the U.S. as cruel and brutal, thereby increasing membership and making the civilian population more sympathetic and permissive to terrorist groups. This aspect applies to Yemen—AQAP has used drone strikes as a propaganda item and recruiting tool, arguing that since the drones will likely target someone regardless of their actual affiliation with AQAP, they may as well join AQAP to exact revenge. There are a number of additional reasons to suspect the risk of blowback may be uniquely high in Yemen. First, as Johnsen argues, AQAP is woven into the social fabric of Yemen in a way that AQ was not in Pakistan. Second, as Johnsen and Hellmich indicate, because AQAP is almost entirely Yemeni (as opposed to AQ’s status as Arab outsiders in Pakistan), it is uniquely hard to distinguish between civilians, sympathizers, and active group members. As a result, the risk of civilian casualties from drone strikes in Yemen may be noticeably higher. Third, the United States’ heavy use of signature strikes and relatively less reliable intelligence informing drone strikes in Yemen may increase the chance of inadvertent civilian casualties. The U.S. lacks reliable on the ground intelligence in Yemen, partially because of the 2015 withdrawal of U.S. special operations forces and evacuation of all embassy personnel as the civil war unfolded.

Finally, it is important to acknowledge what scholars have termed the “balloon effect.” As Johnston and Sarbahi outline, drone strikes may result in terrorist violence simply shifting to different locations, as terrorists adapt and move to geographically secure areas (forests, mountains, etc) or urban areas, where they can camouflage with the local population. This would make sense in Yemen, given AQAP’s history of retreating from battles it perceives it cannot win and simply disperse to other strongholds. However, it is also possible that drone strikes, in significantly restricting the mobility and operational capacity of terrorist groups per disruption/anticipatory effects, may make it more difficult for terrorists to move to alternate locations. Both Johnston and Sarbahi and Mir and Moore examined whether drone strikes shifted violence elsewhere and found that such a shift did not occur. Regrettably, this study does not empirically examine whether drone strikes against AQAP resulted in displacement due to limitations in the location data available constraining the ability of the model to test displacement hypotheses.
Data and Methodology

This study covers the period between December 2009 until January 2018. Data is acquired from a variety of sources. Most of the data on terrorist attacks by AQAP in Yemen is taken from the Global Terrorism Database (GTD), which has a comprehensive dataset of terrorist attacks committed across the globe, compiled from open sources. Unfortunately, the GTD’s final year of data collection was 2017, due to funding issues. As for the 2018 data on AQAP attacks, this study instead utilized the Armed Conflict Location & Event Data Project (ACLED), which gathers data about multiple forms of political violence across Africa, the Middle East, Europe, Latin America, and Asia. Both GTD and ACLED utilize multiple media sources to code relevant information such as location, group responsible, fatalities, etc. GTD’s inclusion criteria require attacks to be intentional, violent (against persons or property), and committed by a sub-national actor which meets their definition of a terrorist group (which AQAP does). ACLED’s dataset for political violence in Yemen includes all violence committed by every identifiable actor in the conflict from 2016 to the present day, including all attacks committed by AQAP. To simplify, GTD counts all instances of violence committed by a terrorist group, while ACLED counts all instances of political violence in Yemen. Thus, although combining the data from GTD and ACLED is an imperfect workaround to GTD’s incomplete 2018 data, there are functionally no cases in which one methodology would “count” an instance of AQAP violence, but the other would not. This is because both datasets count all instances of violence—GTD simply restricts the field of actors to terrorist groups. Because this study solely seeks to examine the effect of the U.S. drone campaign on AQAP, data on terrorist attacks in Yemen committed by non-AQAP groups are excluded. Additionally, a cursory examination of the data indicates that the overwhelming majority of the terrorist attacks committed in Yemen during the study period were committed by AQAP. This lines up with qualitative evidence which indicates that the Islamic State largely failed to establish a meaningful presence in Yemen or hold territory. Specifically, the AQAP outcompeted the Islamic State in Yemen (ISY) because its territorial control made it a materially more attractive option, it was less brutal and presented a more “acceptable face of jihad,” and was more integrated with local tribes.

Two detailed data sources on drone strikes exist. The New America Foundation (NAF) and Bureau of Investigative Journalism (BIJ) both maintain datasets on the U.S. drone war, mostly compiled from verified open source media. Because the U.S. government rarely confirms drone strikes and there is little public knowledge, the data is likely imperfect. Accordingly, this study only includes drone strikes that appear in both the NAF and BIJ datasets, consistent with Manna’s study. An exception to this rule is made for cases in which the NAF database includes data about strikes that were confirmed by U.S. Central Command (CENTCOM) in emails to the NAF or public press conferences. While these strikes largely do not appear in the BIJ dataset, it the data provided by CENTCOM is nevertheless reliable. Because this study examines AQAP specifically, drone strikes against unknown or non-AQAP groups (for example, ISY) are excluded from the dataset. Because Ansar al-Sharia is closely affiliated with AQAP and was created by AQAP during the Arab Spring, strikes against Ansar al-Sharia are included. However, these exclusions are largely irrelevant—almost every drone strike logged in the NAF and BIJ datasets was targeted against AQAP; very few were against Ansar al-Sharia or ISY.

It is worth briefly noting that because not every governorate in Yemen has experienced an AQAP attack or a drone strike, this study only includes data from governorates in which at least one attack or drone strike has occurred—other governorates have no data relevant to the research question. Specifically, this study examines data from Abyan, Ad Dali, Aden, Al Bayda, Al Hudaydah, Al Jawf, Amanat Al Asimah, Dhamar, Hadramawt, Ibb, Lahij, Ma’rib, Saada, Sana’a, Shabwah, and Taiz. Most of the drone strikes and attacks are concentrated in Hadramawt, Abyan, Al Bayda, Shabwah, and Ma’rib. In total, this study includes 1760 observations.
Figure 1: Drone Strike Frequency Distribution in Yemen

Figure 2: AQAP Attack Frequency Distribution in Yemen
This study uses two statistical methods to test the effect of the U.S. drone campaign on AQAP activity. First, a fixed effects model is used to examine the short-run effects of drone strikes on AQAP activity. Fixed effects regression is an econometric model used for panel data. Data is sorted by a cross sectional index variable and a time variable. In this study, the index variable is the governorate in which a drone strike or AQAP attack occurred, and months are used as the time unit. Fixed effects regression analyzes the relationship between the independent variables and dependent variable within each cross-sectional index, and thereby controls for differences across each group that remain relatively constant over time. Thus, the fixed effects model controls for socio-economic and demographic differences (poverty, education, religious and ethnic makeup, etc.) between governorates that may play a role in contributing to terrorism but remain relatively stable over time within each governorate. The time unit sorts the data to examine the short-run consequences of drone strikes, namely statistically significant increase or decrease in AQAP attacks. Including time as a fixed effect also controls for unobserved temporal factors that determine when drone strikes occur, such as changes in weather, the availability of drones with lethal capabilities, bureaucratic and logistical hurdles, and other unobserved effects.

While Johnston and Sarbahi used agency-week for their study and Mir and Moore used tehsil-month as their unit of analysis, this study uses governorate-month. The differences are important to tease out—Johnston and Sarbahi’s agency unit referred to the seven agencies within the FATA region of Pakistan in which the drone campaign was focused, while Mir and Moore used tehsils, a sub-district within each agency, to test the effects of drone strikes in the agency which they were most concentrated (North Waziristan) against tehsils in other agencies. This study’s choice to use governorate-month is driven by three factors.

First, data on drone strikes in Yemen is not granular enough—while older data included the town or additional location data, the majority of data from recent years has only included the governorate in which a strike occurred, which makes subdividing location data into smaller units impossible. Second, Yemen is not Pakistan—AQAP and the drone campaign are spread across multiple regions, rather than just one (FATA in Pakistan). Additionally, there are only 240 total drone strikes across all governorates in Yemen, so simply is not enough data in any single governorate to use Mir and Moore’s procedure and only look at one region without having too small an N. Third, because this study investigates a longer time period than Johnston and Sarbahi’s and covers multiple regions, using weeks as the time-unit is impractical, insofar as it explodes the total number of observations in which no strikes or attacks occur.

The equation for the fixed effects model is given by:

\[ y_{it} = \beta_0 + \alpha_i + \beta_{xit} + \epsilon_{it} \]

In this equation, \( y \) measures terrorist attacks, \( x \) measures drone strikes, \( z \) measures other control variables (explained below), \( \alpha \) measures the unobserved agency effects, \( \epsilon \) signifies the error term. The \( i \) refers to the cross-sectional governorate effects, while \( t \) represents the time index (months).

As explained later, this study examines both the frequency and lethality of AQAP attacks as two distinct dependent variables. In this equation, \( y \) represents the dependent variable, so for the version of the model examining attack frequency, \( y \) represents the number of attacks in a specific governorate (i) in a given month (t). For the model examining lethality, \( y \) represents the number of people killed by all the attacks that occur in a specific governorate (i) within a given month (t). Both of these numbers are acquired from the GTD/ACLED data, which track both the number of attacks and the number of casualties. The \( x \) variable represents the number of drone strikes in a specific governorate (i) in month t, which is sourced from the combined dataset of NAF and BIJ. The agency and time effects (\( \alpha \) and \( \epsilon \)) represent the fixed effects of a given governorate or month. They do not indicate particular data, but are conceptual representations of the way the model accounts for differences in relevant socioeconomic factors between governorates or weather conditions in different months by analyzing the effects of the various independent variables within specific governorates/months.

Utilization of fixed effects model in this study controls for a number of potentially relevant factors that influence AQAP’s willingness to use force against...
civilians. These controls take the form of dummy variables (with one exception), as detailed data is largely unavailable or not relevant to the control variable. A notable limitation is that this study is unable to directly test whether civilian casualties and the success of drone strikes in killing AQAP members affects levels of violence. First, as noted above, data on civilian casualties is notoriously unreliable, often based on anonymous reports or a variety of biased local and western sources. Additionally, Hellmich’s explanation of the unique difficulties in determining who counts as a civilian in Yemen makes numbers even less reliable. More importantly, a significant portion of the data from the Trump presidency on drone strikes is obtained from CENTCOM, which did not provide any casualty numbers for militants or civilians. Excluding this data from the study period is potentially damaging, as an important question posed by this study is whether President Trump’s escalation of counterterrorism operations in Yemen was the causal factor that drone AQAP largely underground, per Kendall’s argument.

The first control variable included in the model is intended to examine the effect of the Trump presidency on the drone war. In March and April of 2017, President Trump launched an unprecedented number of drone strikes in Yemen. Neither the NAF nor BIJ dataset have individualized data points for each strike that occurred in those months that reflect the real number of strikes that occurred, nor do they have data on which regions the strikes occurred in. However, the Pentagon confirmed to the BIJ that forty strikes took place across the country in March, and another twenty strikes occurred in early April, which stands in stark contrast to the average of two to three strikes per month prior to Trump’s presidency. Thus, a dummy variable is coded as being one starting in March 2017. This variable is mostly meant to test Kendall’s qualitative argument that President Trump’s drone surge caused AQAP to go underground, which would be an important result if validated by the data.

Additionally, because the surge began almost as soon as President Trump took office, this variable is also used to account for the general effect of Trump’s presidency on U.S. counterterrorism in Yemen. First, President Trump has notably rolled back a number of bureaucratic checks on the drone program put into place by President Obama, giving the CIA more latitude to operate with impunity, and generally reducing restraint. As a result, civilian casualties under President Trump seemed to climb significantly. This would suggest that drone strikes under President Trump have become less targeted and more prone to result in blowback. Second, non-drone military operations have increased significantly under President Trump, largely as a result of his authorization of generals to make decisions independently. Thus, it is possible that Trump’s presidency will have the opposite effect and see reductions in terrorist activity due to increased non-drone counterterrorism operations. To capture these differing effects, an interaction variable is coded by multiplying the Trump presidency dummy variable with the drone strikes variable such that the interaction variable only outputs a non-zero value in months where strikes occur. This interaction variable isolates for the effect of President Trump’s changes to the drone program (rolling back bureaucratic checks, the lasting effects of the March and April drone surge, etc.), while the original dummy variable captures the overall impact of the Trump presidency on AQAP.

The second set of control variables concerns AQAP’s leadership. As noted above, the vast majority of drone strikes in Yemen are not decapitation strikes—it is rare that drones eliminate an important leadership figure in AQAP. Thus, unlike other studies which test the effect of decapitation strikes as a subset of drone strikes, this study simply codes dummy variables for whether particular leadership figures in AQAP have been killed yet. A zero indicates that they are still alive, while a one indicates that they have been killed. Specifically, this study codes variables for the drone-induced deaths of Nasir al-Wuhayshi (Emir of AQAP, killed June 2015), Anwar al-Awlaki and Samir Khan (notable propagandists/recruiters, killed October 2011), Ibrahim al-Asiri (AQAP’s chief bomb technician, killed in the second half of 2017), Said Ali al-Shihri (deputy Emir, incapacitated December 2012), Jalal Bala’idi (senior commander, killed February 2016), Ibrahim al-Rubaysh and Nasser bin Ali al-Ansi (mufti, or legal expert, and deputy general manager of Al-Qaeda respectively, killed April 2015), Harith bin Ghazi al-Nadhari (senior Sharia official, killed January 2015), Fahd al-Quso (operational commander, killed May 2012), Shawki al-Badani (operational commander, killed November 2014), and Othman al-Ghamdi (operational commander, killed February 2015).
Because al-Asiri’s death date within the second half of 2017 is unknown, he is assumed dead in October 2017, the halfway point of the final six months of the year. While al-Shihri was not technically killed in 2012, the December strike reportedly left him severely wounded and he was killed by a second drone strike in the first half of 2013. This study codes al-Shihri’s “death” as December 2012, as the first strike functionally removed him from the battlefield. If the coefficients for most of these dummy variables trend one way or another and are statistically significant, the study will have confirmed either the hypothesis of Johnston and Price or that of Jordan, Abrahms and Potter, and Abrahms and Mierau.

The third set of control variables concerns various fluctuations in AQAP’s territorial control, grand strategy, and the balance of power over time. Increased territorial control may have a positive or negative effect on terrorism against civilians. As AQAP gains territory, it likely has more freedom to move and operate, as well as greater security for its operatives, and therefore greater ability to stage attacks. Basically, gains in territory reflect favorable shifts in the balance of power between AQAP and its adversaries, and increased activity is expected as a result. On the other hand, greater territorial control may come with added responsibilities. Terrorism may be a tactic to take control of territory, but once territory is acquired, violence against civilians both becomes unnecessary and declines as AQAP attempts to maintain legitimacy in the area it controls. It is also important to note that the effect of holding territory may change over time. As explained above, AQAP’s grand strategy shifted after they lost the territory they gained during the Arab Spring, when they learned that repressive rule was counterproductive and instead shifted to a strategy focused on governance and tribal integration and cooption. Thus, it may be the case that AQAP’s territorial control in the aftermath of the Arab Spring resulted in greater attacks, while the need to maintain legitimacy and governance capacity resulted in decreased attacks during the Yemeni Civil War.

A set of dummy variables attempt to capture these relationships. One variable captures AQAP’s territorial control during the aftermath of the Arab Spring, coded as a one between June 2011—the beginning of territorial control—and June 2012, when tribal militias kicked AQAP out of their territory in Zinjabar and other areas. Another variable captures the subsequent shift in grand strategy towards governance, coded as one for August 2012 onward, the date that al-Wuhayshi sent a letter to his counterpart in AQIM urging a shift to governance. A third variable attempts to capture AQAP’s territorial control during the Yemeni civil war, coded as two starting in February 2015 (immediately following Hadi’s resignation), shifting down to one in April 2016 to coincide with the beginning of the UAE’s attempt to take back territory from AQAP in coordination with the UAE-backed Southern Transitional Council (STC), and shifting finally to zero in February 2018 to coincide with the renewed UAE-STC counteroffensive that seems to have recaptured most of AQAP’s territory. A fourth variable is coded as one starting in April 2015 to represent the beginning of the Saudi air campaign against the Houthis. This variable is meant to capture a potential shift in the balance of power—if the Saudis weakened the Houthis (AQAP’s enemies), AQAP would presumably have more operational freedom. This is backed by qualitative accounts of the Saudi air campaign forcing the Houthis out of territory, leaving a power vacuum that AQAP quickly filled.

The second model this study uses is a difference-in-difference model, as deployed by Mir and Moore and Manna in their studies. The difference-in-difference model seeks to examine the aggregate effect of the drone program by examining trends in terrorist violence before and after drone strikes begin in a given region. While fixed effects measure the short-run effects of individual drone strikes, the difference-in-difference model elucidates whether aggregate terrorist activity increased or decreased as a result of drone strikes beginning in a particular area.

The equation for the difference-in-difference model is given by:

\[
y_{it} = \beta_0 + \alpha_i + \beta(After_{it} \times Treated_{i}) + h_t + \varepsilon_{it}
\]

In this equation, \( y \) represents terrorist attacks, \( \alpha \) measures agency fixed effects, \( h \) measures time fixed effects, \( \beta_0 \) signifies the constant, and \( \varepsilon \) signifies the error term. \( i \) refers to the cross-sectional governorate effects, while \( t \) represents the time index (months). Drone strikes are represented by \( After_{it} \times Treated_{i} \) in which \( After \) is a dummy variable that represents the
month and governorate in which drone strikes begin—if it is in or after the month in which strikes begin, the variable is 1. The Treatedi variable indicates whether a particular governorate was “treated” with drone strikes or not—treated regions are coded as ones, while control regions are coded as zeroes. As with the fixed effects model, y represents both the frequency and lethality dependent variables. It is worth noting that the difference-in-difference model still accounts for the fixed effects of months and governorates. However, the model is noticeably different from the fixed effects model insofar as it is measuring the aggregate effect of drone strikes on the aggregate level of terrorism before and after strikes begin within each region, rather than gauging the effect of single strikes on terrorist activity.

There are some coding complexities in the difference-in-difference model. First, a select few governorates have very limited drone strikes. This study only codes a governorate as having been “treated” with drone strikes if the number of strikes equals or exceeds five over the time period studied, as a single strike is unfit to constitute a significant “treatment.” Excluded therefore are Dhamar, Saada, Sana’a, and Lahij, which received one to two strikes, and included is Al Jawf, which received five strikes. Second, in Abyan and Shabwah, the first drone strike occurs in the first month of the study period. Because the difference-in-difference model needs to capture a before-and-after picture to examine the effect of treatment, these strikes are excluded and coded as zero, and treatment is considered to have begun after the second strikes in Abyan and Shabwah (both of which occur in 2011). Third, it is difficult to pinpoint the time at which the program “started” in control regions that had received no strikes. In Mir and Moore’s study, only one region is treated, meaning that the start date for the drone program is a single value for all regions. However, in this study, drone strikes begin at different times in different governorates. Yet, the difference-in-difference model requires a start date for treatment in the control groups which are not treated with drone strikes. This is because the model seeks to examine the difference between the pre- and post-treatment dependent variable for the treated group, the difference between the pre- and post-treatment dependent variable for the control group, and the difference between those differences. This study codes the beginning of drone treatment for control governorates as May 2012, the latest start date for drone strikes in any of the treated regions. This choice to skew towards a later start time is intended to ensure that there is sufficient data constituting the before-treatment period. Because the study period spans from December 2009 to January 2019, a treatment start date is chosen as close to the midpoint of the data as can be allowed.

To code each variable, the drone strike independent variable from the fixed effects model is converted into the dummy Afterit variable—the first drone strike in a treated governorate marks the shift from zero to one. For control governorates, the shift from zero to one occurs in May 2012, as explained above. The Treatedi variable is a dummy variable to indicate whether a given governorate is a control or treated governorate—treated governorates are coded as one, controls as zero. Thus, the Afterit*Treatedi variable is obtained by multiplying the two together, creating an interaction variable that describes whether being a treated governorate (Treatedi) changes the before and after effects of the drone program beginning (Afterit), thereby uncovering the effects of treatment with drone strikes.

Results and Discussion

This study ran the fixed effects model repeatedly with slight alterations to establish any trends that persisted across different iterations of the model. In the first iteration, the model was run with every control variable and no alterations. In the second iteration, control variables examining the effect of decapitation are removed. In the third iteration, the territorial control variables (Arab Spring, Yemeni Civil War, Saudi Strikes) are omitted, while the decapitation variables remain. In the fourth iteration, both territorial control variables and decapitation control variables are omitted. In the fifth iteration, only the interaction variable is included, along with the drone strikes independent variable. In the sixth iteration, all control variables are omitted, and only the drone strikes independent variable remains.

These iterations were used for two reasons. First, omitting certain control variables tests to see if the primary independent variable being studied (drone strikes) has a consistent coefficient in each iteration of the model. Ideally, drone strikes will have a consistently positive or negative coefficient that remains
statistically significant in most, if not all iterations of the model. Second, omitting certain control variables reduces the chance of accidentally finding certain control variables to be statistically significant simply due to the sheer number of controls. Particularly in the case of the decapitation variables, where ten variables are coded, it is important to ensure results are not being inadvertently p-hacked.

In addition, each iteration of the model was run eight times. There are two dependent variables being tested—terrorist attacks and the lethality of attacks. However, in line with Johnston and Sarbahi, this study codes two new dependent variables to normalize measures of terrorist attacks and casualties to capture the number of attacks and casualties per thousand residents of the given governorate. Thus, each combination of independent variables is run on four slightly different dependent variables—attacks, casualties, population adjusted attacks, and population adjusted casualties. Finally, each run of the model is duplicated with standard robustness checks to check for heteroskedasticity and to determine the strength of the results. The tables below display the outputs for each iteration of the fixed-effect model that included the robustness checks and heteroskedasticity correction. The results of the model without robustness checks can be found in the appendix.

The results of the fixed effects model are relatively consistent. The drone strikes independent variable had a positive coefficient in all 48 runs of the model, and the coefficient remained roughly the same value within each the four dependent variables’ different iterations. The coefficient’s value obviously changed between the different dependent variables, as the attack and casualty numbers were not identical, and the attack variable was not numerically identical to the population adjusted attack variable. In the first iteration of the model, drone strikes have a statistically significant p-value (p<.05) across all four dependent variables. This persists across each subsequent iteration of the model, with drone strikes generally obtaining a lower p-value in runs of the model with fewer control variables (the iterations without the decapitation variables). Thus, drone strikes tended to have a counter-productive effect, causing increased attacks and increased casualties.

Interestingly, the interaction variable tracking President Trump’s effect on the drone campaign has a negative coefficient across nearly every run of the model. This suggests that President Trump’s drone surge in March and April of 2017 and his overall reduction in restraint in the use of drones may have slightly reduced attacks and lethality, as posited by Kendall. However, while the Trump presidency interaction variable is statistically significant prior to running robustness checks, the significance almost completely vanishes in every run of the model with robustness checks. As a result, it is more likely than not that Trump’s presidency has not had any significant effect on AQAP.

Unfortunately, the results of the fixed-effect model concerning drones are not strong enough to form definitive conclusions. In nearly all of the runs of the model with standard robustness checks, both drone strikes and the interaction variable lose statistical significance, though they sometimes obtain p-values below .10. While the coefficients retain their values (positive for drone strikes, negative for the interaction variable), the p-value is not low enough to reject the null hypothesis. Thus, these results suggest that drones may have no significant effect on terrorism in the short-run. However, it may be the case that the robustness check eliminated significance for the drone and interaction variables simply because they varied sporadically over time, and strikes did not occur frequently enough to definitively pass the robustness check. In contrast, control variables were largely unaffected by the robustness check on account of being dummy variables that rarely changed.

The decapitation control variables rarely obtain statistical significance. In some runs of the model, the variables tracking the deaths of Anwar al-Awlaki/Samir Khan and Ibrahim al-Asiri obtain statistical significance. However, this seems likely due to random chance and the simple fact that a large number of decapitation dummy variables are included in the model. If decapitation had a noticeable effect on AQAP violence, we would expect all, or almost all of the coefficients to have a consistently positive or negative value. This is not the case. The decapitation variables are roughly evenly split between positive and negative coefficients, suggesting decapitation has no discernable effect on violence. Of course, this does not necessarily justify the conclusion that decapitation is totally ineffective—it could be the case that there are unobserved longer-term consequences of decapitation that are not
captured by the data. Perhaps the elimination of much of AQAP’s leadership in the first half of 2015 gradually degraded the group’s capabilities, and played a role in the group going underground in late 2018. It could also be that this delay was due to AQAP’s continued hold on territory after the 2015 decapitations. However, further research is needed to examine the role of decapitation in modifying AQAP’s activities. This model’s results seem to suggest that decapitation has been largely irrelevant in the Yemeni context.

The territorial control variables are interesting and somewhat counterintuitive. The variable measuring Saudi strikes is omitted due to exact collinearity with one of the decapitation variables in the first iteration of the model. However, when the decapitation control variables are removed and the variable for Saudi strikes is included, the variable obtains statistical significance with a negative coefficient. The change suggests that the Saudi bombing campaign against the Houthis reduced AQAP violence. This odd result persists even with robustness checks, but is perhaps explained by the other territory variables. If the Saudi variable is omitted, the Yemeni Civil War variable obtains statistical significance with a negative coefficient when it otherwise remains insignificant. This implies that insofar as the Saudi air campaign largely overlapped with the Yemeni Civil War, and that the air campaign assisted in AQAP’s territory grabbing, the most probable result was that AQAP’s territorial control during the civil war reduced their propensity to use violence against civilians. As expected, the Arab Spring territory variable is positive and significant when decapitation variables are excluded from the model, suggesting that AQAP’s control of territory prior to the Yemeni civil war aided them in their repressive rule of the civilian population. This remains the case when running the model with robust standard errors.

The variable that tracks whether the al-Wuhayshi letter had been sent yet is significant and has a positive coefficient. This is unsurprising—if the letter reflected a grand strategy shift away from violence against civilians, a negative coefficient would be expected. This is perhaps explained by the fact that the governance letter was sent in 2013, prior to AQAP’s territory gains from the civil war. All of the qualitative research outlined above indicates that AQAP used a governance-focused grand strategy during the civil war when they controlled territory. However, perhaps the letter only reflected a grand strategy shift that was conditional on AQAP’s territorial control. Between 2013 and 2015 after the letter was sent, AQAP operated as a normal terrorist entity. However, once they controlled territory, the governance-focused grand strategy outlined by al-Wuhayshi’s letter went fully into effect. This explains the reduction in violence outlined by the Yemeni Civil War variable.

The process for running the difference-in-difference model was somewhat similar to the fixed effects model. The model was run in multiple iterations. Because difference-in-difference only considers the effect of treatment by drone strikes in the aggregate, the control variables are not included in the model. Furthermore, the nuance of their effects are more accurately documented by the fixed effects model. The difference-in-difference model is obtained by changing the drone strike data into a series of dummy variables (as described above) and then running the data as a “fixed effects” model. Meanwhile, the difference-in-difference model still controls for fixed time and group effects. The model was first run with the Afterit variable and the Afterit*Treatedi variable on each of the four dependent variables. The model was then run with standard robustness checks for each dependent variable.

In the two studies using the difference-in-difference model, both Manna and Mir and Moore select their study period such that treatment begins at the half-way point. Unfortunately, this study was primarily designed to capture the impact of the drone program on AQAP throughout the changing political conditions in Yemen, with an additional focus on whether Trump’s presidency has noticeably impacted the drone program. As a result, the period under study is quite long, and treatment begins around the one-fourth mark of the entire study period. Because difference-in-difference is a before-and-after examination of the effects of treatment, this creates the concerning possibility that data from the before-treatment period (December 2009 to 2011-2012) will be given significantly more weight than the after-treatment period. Because there are fewer observations in the “before” period, any outliers in the “before”-treatment data on terrorism will be amplified, and there is less data to establish a clear trend. To correct for this possibility, the iterative process described above is repeated, ex-
cept data from after September 2014 is excluded. This results in the treatment period beginning at around the half-way mark (May 2012) for most of the governorates, for a total of 928 observations. Thus, the model is run a total of eight to sixteen times with the full data, and eight times with the shortened study period.

The results of the model are displayed below. As with the fixed effects model, only the heteroskedasticity adjusted, robustness checked runs of the model are displayed, with full results in the appendix. Graphs of the data are also included. The blue spike represents the beginning of the treatment period, and the orange line tracks the number of attacks in a given month. The x-axis converts months into numbers, and there are 110 months total within the study period. Recall that while all regions have a “beginning” date for treatment, only Abyan, Al Bayda, Al Jawf, Hadramawt, Ma’rib, and Shabwah were actually treated with drone strikes. The results from the difference-in-difference model suggest that drones had a counterproductive effect. When using the full data, the model demonstrates that governorates treated with drones saw an increase in attacks (a positive coefficient). The p-value retains statistical significance below .05 after standard robustness checks for both the nominal attacks variable and the population-adjusted attacks variable. Interestingly though, in the runs of the model testing the casualties-dependent variable which included the full study period, while the coefficient for the relevant variable (Afterit*Treatedi) remains positive throughout, it never obtains significance, even at the .10 level. Because we would roughly expect casualties to climb as the number of attacks increased, this result tentatively suggests that although drone strikes increased the frequency of attacks, the lethality of attacks did not increase at the rate we might expect with increased attacks. This may mean that more attacks occurred, but those attacks were slightly less lethal relative to attacks prior to treatment (which explains why the lethality variable was still positive, but did not obtain significance).

In the runs of the model with data removed, the results remain largely the same. For the attack dependent variables, the Afterit*Treatedi variable has a positive and statistically significant coefficient, except for the run of the model for the non-adjusted attacks variable using standard robustness checks, in which the variable is only significant at the .10 level, rather than the ideal .05. Interestingly, the casualty-depen-

Death From Above
and result in groups cracking down on civilian populations suspected of having become informants. Yet, the lack of statistical significance when the model is run with standard robustness checks suggests that none of these explanations can be strongly confirmed by this study.

Third, the long-run effects of the drone program are in some ways clear, and in other ways not. If Mir and Moore's argument about the anticipatory effects of drone strikes were correct, the difference-in-difference model (and perhaps the fixed effects model as well) would demonstrate that the beginning of treatment via drone strikes reduced terrorist violence. This appears to not be the case. However, a few caveats are important. First, the results concerning casualties for the full dataset can be interpreted to support the anticipatory effects argument. If, for example, drone strikes disrupted AQAP's ability to train recruits and made organization more difficult, it might be expected that even if attacks increased due to recruitment-related blowback, the lethality of those attacks would decline (as operatives would be less lethal and less well trained). However, this reading of the data is likely incorrect, as the coefficient for the runs of the model examining lethality remains positive, and the version of the model that excludes post-September 2014 data indicates that drone strikes resulted in greater lethality. If drones did reduce attack lethality, the coefficient would be negative and significant, but instead, it is positive and not significant, suggesting strikes had little effect. Second, it is also possible to read the results for the Trump presidency interaction variable as partial confirmation of the anticipatory effects argument. The fact that the interaction variable obtains significance with a negative coefficient tentatively suggests that Kendall may be somewhat correct in assessing that President Trump's ramping up of drone strikes played a role in pushing AQAP underground. However, the fact that both Trump presidency variables lose all statistical significance when robustness checks are included suggests Trump's presidency has had little meaningful effect.

Despite these caveats, it is important to stress that this is an exceedingly generous interpretation of the results from each model. By and large, drone strikes appear to be positively related with terrorist violence in both models, but particularly in the difference-in-difference model. The fact that the difference-in-difference model obtains statistically significant results after standard robustness checks is important. Adding to the finding is that the coefficients almost universally indicate the drone program has, in the aggregate, increased terrorism. Put simply, this result may be the first empirical confirmation of the blowback hypothesis in Yemen—drone strikes are counterproductive in the long run. For the many reasons outlined above, Yemen appears to be uniquely at risk of greater blowback via increased recruitment compared to Pakistan.

Given that recruitment is likely not an immediate effect of a drone strike (radicalization and training to conduct attacks likely require significant time), we would expect the recruitment arguments to be confirmed if drone strikes appeared to increase terrorism in the long run. This appears to be the case, based on the results of the difference-in-difference model. To be clear, it is impossible to directly attribute the results of the difference-in-difference model to increased recruitment, as recruitment would only indirectly increase the number and lethality of attacks, and the data do not indicate specifically why terrorism increased. However, among the causal explanations presented to support H2 (that drone strikes increase terrorism), recruitment seems to fit best as an explanation for drone strikes increasing terrorist violence in the long run. Other explanations for increases in violence such as decapitation, retaliation, and crackdowns on civilians seem more likely to reveal themselves as short-run effects of drone strikes. The fixed effects model indicates that these effects, while they may have been present in Yemen, are not robustly confirmed by the data. However, the strength of the results from the difference-in-difference model suggests that some other explanation (blowback) accounts for the aggregate increase in violence as a result of the drone program.

Conclusion

This study has examined the effects of the U.S. drone program on AQAP's activities in Yemen. After a detailed examination of the qualitative evidence on Yemen, the bulk of the research seemed to indicate that drone strikes would have generally counterproductive effects, though some outliers suggested that President Trump's surge of drone strikes in 2017 may have played an important role in pushing AQAP underground, and
much of the empirical work on the U.S. drone program in Pakistan has found that drones were effective at reducing terrorist violence. This study’s results suggest that the drone campaign has had counterproductive effects in Yemen. In the fixed effects model, albeit somewhat weak results suggest that 1, drone strikes may increase attacks and casualties in the short run; 2, President Trump’s surge in drone strikes has not had a significant effect on AQAP violence; 3, decapitation strikes have no discernible effect on violence; and 4, AQAP’s grand strategy shift between the Arab Spring and Yemeni Civil War did result in the expected violence reduction. The difference-in-difference model paints a clearer (and statistically significant) picture—drone strikes increased terrorist attacks (most likely due to blowback), and likely increased casualties as well. Across all four versions of the dependent variable, drone strikes were positively correlated with increased attacks and increased casualties, suggesting that no matter how the data is sliced, drone strikes are a poor counterterrorism tool. Even if drones are occasionally “effective” at killing terrorists, these results indicate that the harm outweighs the benefit. For a counterterrorism tool to be effective, it should fulfill the ultimate goal of reducing terrorism. By this metric, drone strikes have been a counterproductive failure in Yemen. However, further research is still needed. The empirical effects of the U.S. drone program remain under-researched, particularly in countries for which data are available (Yemen, Somalia, Afghanistan), but for which research has largely not yet been conducted. This study has attempted to partially fill a gap in the literature surrounding Yemen, and seems to support the qualitative arguments indicating that there is a greater risk of blowback to the U.S. drone program in Yemen than in Pakistan. However, further work is needed to illuminate the ongoing effects of President Trump’s changes to the drone program, as well as more carefully disaggregate the effects of drones on AQAP from the political conditions in Yemen shaping AQAP’s ability to conduct attacks.

Appendix

Due to space limitations, the author has chosen to upload the data used by this study and the full results generated by each iteration of both models to be publicly accessible for others to view and use.

Files are available at the following URL: https://drive.google.com/drive/folders/13Cs1U-kFwCTjV960w4LY-98tuQWHigxS6?usp=sharing

Endnotes

4. Sergie, “AQAP”.
8. Sergie, “AQAP”.
12. Sergie, “AQAP”.
13. Sergie, “AQAP”.
17. Sergie, “AQAP”.
4967.2012.00554.x.
82. Walsh, “Precision Weapons,” 519-20.
91. Johnsen, “How We Lost Yemen.”
92. Johnsen, “How We Lost Yemen.”
116. David Knoll, “Al Qaeda in the Arabian Peninsula


161. Johnsen, “How We Lost Yemen.”

162. Johnsen, “How We Lost Yemen.”


Bibliography


Page, Michael, Lara Challita, and Alistair Harris. “Al-Qaeda in the Arabian Peninsula: Framing


Beyond *Mano Dura*: An Integrated Strategy for the Northern Triangle

Ryan Witmer Keating, Elliott School of International Affairs, George Washington University

Abstract

The violence that plagues the Northern Triangle today is a function of underlying sociological factors. Governmental efforts to address this violence and its contributing factors have fallen short. Drawing on relevant literature the author breaks down why these policies have fallen short in reducing violence. Based on this analysis, the author will use the lessons of the counterinsurgency (COIN) framework and the case study of Colombia to outline an integrated security strategy (ISS) for the countries of the Northern Triangle.

Section 1: Introduction

Violence, criminality, and gang membership have made the Northern Triangle one of the most dangerous areas of the world. El Salvador, Guatemala, and Honduras had an average rate of over forty-three intentional homicides per 100,000 citizens in 2017. Even acknowledging the improvement in 2017, Figure 1 illustrates that this is dramatically higher than other countries in the region.\(^1\) Criminality also expands beyond homicides with an average of 24 percent of the population reportedly being victimized by crime in 2017.\(^2\) Gangs play a significant role in propagating this violence and criminality both because their...
members number between an estimated 54,000 and 85,000 across the countries and because of the growing sophistication of these criminal organizations.\(^3\) As three relatively unconsolidated democracies with weak governmental institutions, El Salvador, Guatemala, and Honduras have struggled to respond to these challenges. The inability of governments to respond sufficiently to violence and criminality has stifled economic growth and significantly increased the desire of citizens to emigrate above the regional average.\(^4\) This poses a serious threat to these young democracies because insecurity continues to be one of the most likely motivators of a military coup.\(^5\) Given these long-term threats, this paper will advocate for the design and implementation of an integrated security strategy (ISS) for successfully mitigating both the threat of the gangs and the underlying drivers of violence and criminality.\(^6\)

Beginning with a review of relevant literature, the paper will first examine Durkheim and Gurr’s work on how sociological factors, including sociopolitical exclusion, lack of economic opportunities, ineffective governance, and cycles of violence act as drivers of violence and criminality. This section will also look at the development of gangs from unorganized street gangs to sophisticated transnational criminal organizations with a focus on the similarities between third-generation gangs and insurgent actors. From there, the literature review will also examine Colombia’s counterinsurgency campaign to highlight the value of the Counterinsurgency (COIN) framework and the Problem-Driven Iterative Adaptation (PDIA) approach for addressing both the issue of the gangs and the deeper sociological drivers of violence and criminality.\(^7\)

The author will then examine the dominant policy approaches pursued by the countries of the Northern Triangle thus far. This section will break down those approaches into three types of policies: Mano Dura, Mano Amiga, and La Tregua. This section will explain the rationale behind these three policy approaches, give examples of the implementation of these policies in different countries and assess the impact of these policies. By examining where these policies were successful and where they fell short, this section will derive lessons to inform future policy approaches.

Beginning with a review of relevant literature, the paper will first examine Durkheim and Gurr’s work on how sociological factors, including sociopolitical exclusion, lack of economic opportunities, ineffective governance, and cycles of violence act as drivers of violence and criminality. This section will also look at the development of gangs from unorganized street gangs to sophisticated transnational criminal organizations with a focus on the similarities between third-generation gangs and insurgent actors. From there, the literature review will also examine Colombia’s counterinsurgency campaign to highlight the value of the Counterinsurgency (COIN) framework and the Problem-Driven Iterative Adaptation (PDIA) approach for addressing both the issue of the gangs and the deeper sociological drivers of violence and criminality.\(^7\)

The author will then examine the dominant policy approaches pursued by the countries of the Northern Triangle thus far. This section will break down those approaches into three types of policies: Mano Dura, Mano Amiga, and La Tregua. This section will explain the rationale behind these three policy approaches, give examples of the implementation of these policies in different countries and assess the impact of these policies. By examining where these policies were successful and where they fell short, this section will derive lessons to inform future policy approaches.

In the final major section, this paper will outline the basic objectives and framework for a future ISS for the Northern triangle. This ISS will consist of four lines of effort to address both the direct threat posed by gangs as well as the underlying sociological factors driving violence and criminality. These lines of action will include incremental improvements in governance, security, economic empowerment, and sociopolitical inclusion. This section will draw heavily on Colombia’s 2002 Democratic Security and Defense Policy (DSDP) to give examples of what these policies would look like.
Section 2: Literature Review

In order to develop an ISS, three relevant bodies of literature must be examined. The first of these is sociological literature on drivers of violence and criminality. Analysis from Émile Durkheim and Ted Gurr illuminates how sociological conditions fuel violence and criminality. The literature on the organization and operation of gangs, particularly the third generation gangs, builds on this body of knowledge. It illustrates the challenge these more sophisticated criminal organizations pose to the state and state policy and highlights the similarities between third generation gangs and insurgents. The literature on insurgency and counterinsurgency offers valuable insights for understanding the corrosive effect of gangs on the state and the steps that an integrated security strategy must take to address the underlying sociological drivers of violence while marginalizing gangs.

Durkheim’s theory of anomie and Gurr’s theory of relative deprivation provide the framework for tying sociological conditions to violence and criminality in the Northern Triangle. Durkheim posits that “collective order” serves as a regulator of human expectations, which he describes as “passions”, and keeps them from growing beyond the realm of possible attainment. He warns that “sudden depression, sudden prosperity, and rapid technological change” can lead to a breakdown in the social order, and that this “de-regulation or anomie” allows a gap to develop between aspirations and fulfillment. Durkheim argues that this gap, and the frustration it entails, fosters deviant behavior such as crime and violence.

Gurr refines the link between sociopolitical conditions and criminality and describes how a gap between expectations and attainment in realms such as economic attainment and governance can lead to violence with his idea of “relative deprivation.” Gurr describes the idea of “relative deprivation” as the tension between an individual’s actual state and the state the individual believes they deserve. When the gap between people’s expectations and their attainment grows, their frustration with their circumstances creates pressure for deviant behavior. Gurr’s contribution is important in a discussion of the Northern Triangle because he expands beyond just the sociopolitical dynamics discussed by Durkheim and argues that relative deprivation can occur with regard to other expectations such as a deficit of security, economic success, self-actualization, and government services.

Mo Hume directly ties these abstract sociological ideas to the Northern Triangle by looking at how major social and political shifts during the 1990s fostered the development of street gangs in El Salvador. These sociopolitical shifts broke down the collective order that had regulated expectations while also increasing expectations about the dividends peace and democracy would provide beyond what could be reasonably attained. As time passed, democratic reforms stalled, and the economic benefits proved to be less than many had expected (Figure 2). The combination of these factors created pressure for deviant behavior and led to the rise of street gangs. Though Hume does not directly examine Guatemala and Honduras, the basic link he creates between sociological conditions and violence directly applies to these cases as well. Recognizing the impact of sociological conditions on violence and criminality is important because it shows that security policies must address the full spectrum of factors contributing to relative deprivation and the violence that it drives.

Figure 2: Data from World Bank

![2018 GNI per capita (PPP)](image)
Beyond adopting a holistic recognition of the factors driving violence, policymakers must also examine the complicating role of gangs for policymakers attempting to address the underlying sociological factors. The gangs of the Northern Triangle today include a number of third-generation gangs, most notably MS 13 and Barrio 18. Though the literature disagrees to some extent and the differentiation between generations is blurry, most scholars believe that third-generation gangs are distinguishable from less sophisticated organizations by their territorial control, their internal organization, and their calculated use of violence. As a result of their sophistication and ambition, third-generation gangs pose a greater threat to “the modern nation-state and its institutions” than their first- or second-generation counterparts. In terms of organization, third-generation gangs are “complex hierarchical organizations that operate according to a division of labor.” In the case of MS-13, this consists of a three-tiered hierarchy presided over by the Ranfla, a strategic level command structure that sits atop the hierarchical organization. This strict hierarchy is crucial for understanding the operations of third-generation gangs because it allows the leadership of the gangs to react quickly and strategically to employ violence to protect the organization’s lucrative illegal activities. As shown by Figure 3, third-generation gangs are also notable because they will begin to move into and “control ungoverned territory within a nation state and/or begin to acquire political power in poorly governed space.” They do this with the intention of providing security and freedom of movement for their operations. While it is not their objective, if this gradual process of taking power from the state continues, the end result will be “the violent imposition of a radical socioeconomic-political restructuring of … governance in accordance with criminal values.” Even before reaching this level of power, third-generation gangs are detrimental to a state because they normalize violence, legitimize corruption, distort market mechanisms, disrupt equitable commercial transactions, and degrade the environment they operate in. Thus, while third-generation gangs perpetrate violence without the objective of advancing an ideology, their violence is nonetheless corrosive to the authority and legitimacy of the state. Barnes describes this as “criminal politics” because gangs are confronting the state with the objective of accumulating wealth and informal power without aspiring to hold formal political power.

Examinations of gang operations demonstrate that they have organizationally evolved and are growing in their economic, social and political power. In examinations of MS 13 in Honduras, Babineau and Farah illustrate the way the criminal organization wields its military, economic, social and political power. As Figure 4 shows, the organization is expanding its territorial control with the strategic objective of controlling the lucrative trafficking routes. MS 13 also directly interacts with the Honduran political system. It does this by funding the campaigns of mayors and local legislatures friendly to its agenda, charging candidates to campaign in areas under their control and banning politicians it opposes from campaigning in areas under its control. In addition to these interactions with the existing political system, the organization is operating as a counter-state actor in some ways. This includes holding ‘judicial’ hearings to resolve disputes, limiting extortion within the territory they control, creating economic opportunities, and establishing “rudimentary social services.” While these activities can be explained as the organization acting in its own self-interest by accumulating political and economic capital, the provision of social services shows that third-generation gangs are undermining the state in more sophisticated ways.

Although gangs mainly focus on economic rather than political objectives, this component does not disqualify gangs because third-generation gangs are more sophisticated in their interactions with the state. Bailey and Taylor examine the strategies criminal organizations and governments employ to advance their respective interests of “unfettered criminal activity” and “the appearance of order.” The authors conclude that, because of the inherent link between the gang’s economic interests and the exercise of political power, sophisticated criminal organizations have an incentive to leverage violence to advance their political interests. As Barnes points out with his idea of “criminal politics”, conceptualizing criminal violence as neatly separated from political violence is a false model of reality that is “neither conceptually nor theoretically justified.” This is to say, though holding political power and making political decisions is not the objective of gangs, their actions nonetheless generate political capital and they can use that political capital to influence decisions that affect their interests. Bailey
and Taylor’s piece provides a useful contribution to a discussion of gangs in the Northern Triangle because it illustrates the oftentimes unseen political influence gangs leverage for their own benefit and validates their classification as insurgent-like.

Figure 3: Characteristics of gang generations

The expansions of gangs beyond purely economic interests means that gangs no longer fit neatly into the category of criminals. Instead, they act increasingly similarly to insurgents in all but their ultimate objective. The term insurgent conjures images of guerilla fighters, but insurgency is more complex than just non-traditional battle tactics. Many scholars, including Jeremy Black, adopt the dictionary definition of insurgency as “rising in open resistance to established authority.”28 The vague nature of this definition has inherent limitations, but endeavors to encompass the incredibly diverse forms that insurgency has taken around the world and that persist today. Other scholars build on Black’s broad definition by specifying that insurgents are “actors [who use] violence to contest the sovereignty of the established regime with the aim of political change.”29 This paper will utilize this second, more specific, definition of insurgency because of its inclusion of political goals.

Classifying certain gangs as similar to insurgent groups is useful because an extensive body of work already exists on how to address the complex factors underlying the rise of insurgent groups. The end of the Cold War and operations in Iraq and Afghanistan have led to a new focus on the principles of COIN based on the evolving needs of Western powers. The theory of counterinsurgency developed during this period is often referred to as ‘population-centric counterinsurgency’ because of its focus on protecting and providing services to the population to win their support.

This differs from ‘enemy-centric’ counterinsurgency that places an emphasis on finding and eliminating insurgents.

The body of population-centric counterinsurgency literature acknowledges the case-specific nature of counterinsurgency operations but retains three major principles for counterinsurgents across all cases.30 First, the counterinsurgent should seek to accommodate and re-incorporate the reconcilable opposition back into the government and society. It is unquestionably difficult to identify the reconcilable subset of insurgents so the counterinsurgent should achieve this by creating sustainable avenues for insurgents to rejoin society where reasonable and possible. Second, the counterinsurgent should selectively and carefully apply force to the insurgents, particularly the irreconcilable insurgents. This force must be applied in a deliberate and discriminate way to prevent a negative backlash from the populace, but it is nonetheless necessary to sideline the insurgent group through the use of force.31 Finally, and most importantly, the counterinsurgent should win over the population by providing for the population through the provision of public goods.
such as security, education and healthcare. These three principles form the basic framework of the COIN approach.

**Figure 4: Graphic published by Stratfor showing the geographic distribution of homicides in the country**

In order to implement these three principles, advocates of COIN argue for incremental improvements along four parallel lines of effort in four principle areas: governance, security, economic opportunities, and sociopolitical inclusion. Colombia's successful campaign against the FARC demonstrated that incremental improvements in each of these categories—combined with regular evaluations of effectiveness—have demonstrable effects. While the rollout of these parallel lines of effort takes time, its programs and initiatives are able to build upon each other, correct shortfalls and learn from experience.

As the literature stands, several things are clear: A solution will only be sustained if it addresses the previously overlooked sociological challenges in the Northern Triangle to successfully improve conditions in the long term. In addition to this, because of the growing sophistication and influence of third-generation gangs in the Northern Triangle, it is clear any attempt must acknowledge and account for the complex security challenge posed by the gangs and their interactions with the state. Finally, as the literature shows, COIN provides a useful framework for addressing the holistic causes of violence while dealing directly with the threat posed to state power by third-generation gangs.

**Section 3: Previous Policy Approaches**

The Northern Triangle countries have not been idle as violence and criminality grew into a more urgent challenge; however, policy approaches thus far have been insufficient. Similar to the ‘tough on crime policies’ pursued by the United States in the 1980s and 1990s, the first group of policies sought to suppress criminal activity by strengthening security forces and increasing penalties for criminal offenses. First implemented in the Northern Triangle the early 2000s, these types of policies are collectively referred to as *Mano Dura* policies. *Mano Amiga* was a second group of policies that sought to complement the shortfalls of *Mano Dura* by implementing crime prevention, intervention and reintegration programs. The third approach, *La Tregua*, saw the governments in El Salvador and Honduras experimenting with negotiations between the government and gang leaders to reach a compromise and reduce violence. Examining the successes and shortcomings of these previous policies and contrasting them with evidence from Colombia is a worthwhile endeavor because it helps inform how future strategies should be designed.

**3.1 Mano Dura**

*Mano Dura* (literally translated as “strong hand”) refers to a series of policies implemented in the 2000s that responded to rising homicide rates by directly suppressing the most visible forms of criminal activity and gang membership, even where it was less harmful. These policies differed slightly from country to coun-
try, but were characterized by three common features: new laws facilitating arrest based on subjective evidence, the elimination of procedural protections, and the militarization of internal security forces.32 These tough-on-crime measures restricted the legal rights and protections of gang members and were made possible, at least in part, by the sociopolitical othering of gang members and the polarization of domestic politics.33 In Honduras, presidents Ricardo Maduro, Juan Hernández, and others championed laws allowing the military to engage in police activity, increased penalties for attacks on security forces, and created new militarized police units such as the Tropa de Inteligencia y Grupos de Respuesta Especial de Seguridad (TIGRES).34 El Salvador followed suit and passed new laws involving the military in security functions and criminalizing gang membership through Ley Anti-Mara. Ley Anti-Mara would be overturned after being found unconstitutional by the Salvadoran supreme court because it allowed for “arbitrary differentiation for crimes committed by gangs”.35 Nonetheless, President Saca would push forward Plan Super Mano Dura that included a number of similar measures. Even in Guatemala, where the implementation of Mano Dura policies was the most limited, President Jimmy Morales ordered 4,500 military members to deploy in support of police forces in 2016.36

Figure 5: total prison population in the Northern Triangle according to World Prison Brief

Despite the promises of the politicians who enacted them, the Mano Dura policies did not have the intended effect of reducing crime and homicide rates. In El Salvador, the criminalization of gang membership led the police to round up thousands of young men based on little more than their appearance.37 These cases clogged the court system and did little to improve a justice system with a homicide conviction rate of only 5 percent.38 This imbalance in prosecution pushed low level gang members into a gang-run prison system that functioned as a “finishing school” for young gang members who might have otherwise left the gangs.39 Thus, while serious offenders were able to operate with relative impunity, low level offenders were crammed into an overcrowded prison system where they were further indoctrinated in exchange for the protection and benefits afforded by gang membership.40 In addition to this, credible reports of excessive force and human rights violations by the police and the military persist to the present day.41

It is clear the Mano Dura policies have not had the intended effect; conversely, there is the argument to be made that the Mano Dura policies have even had a counterproductive effect in El Salvador.42 In Honduras, where they were reported to be most effective, the Mano Dura policies contributed to slightly reducing the homicide rate. However, this success has been limited by the fact that two thirds of the population of Honduras is unemployed or underemployed.43 Honduras illustrates the crucial point that, even where repressive security measures are most effective, security in isolation will not solve the underlying challenges that persist across the Northern Triangle. Going forward, these types of security measures must be understood not as a solution in their own right, but rather as a tool in support of a broader and more comprehensive approach.

3.2 Mano Amiga

The second major group of policies, collectively referred to as Mano Amiga, were developed in response to the failures of Mano Dura and facilitated by support from the United States, NGOs, and others. These policies moved toward community-oriented policing through crime prevention programs, community interventions, and reintegration measures to rehabilitate criminal offenders. There is strong evidence that these types of security policies are more effective at reducing crime than the suppression strategies employed by Mano Dura.44 As such, they have been experimented with across the Northern Triangle on a small scale in an attempt to address the failures of Mano Dura.45 Honduras has launched programs that have provided
family and individual therapy for youth deemed to be at risk of joining gangs. El Salvador’s País Seguro included prevention and intervention strategies coordinated through a new Secretaria de la Juventud to better integrate government policies. The types of programs promoted by Mano Amiga are promising and have been successful where they have been implemented. A USAID evaluation of CARSI supported programs found that community-based crime preventive measures reduced victimization and improved security across all three countries of the Northern Triangle. However, even with US support, these programs have been chronically under-resourced and under-funded. In the case of El Salvador, LAPOL polling shows that there have been improvements to the level of community engagement by police. However, these expansions are still relatively limited despite the Mano Amiga policies featured prominently in El Salvador Seguro. This is, in no small part, the result of only 6 percent of the funding raised through anti-crime tax being spent on these types of policies with the rest going towards traditional security policies. While the newly created Secretaria de Juventud was a step towards a comprehensive understanding of security, the office lacked the institutional strength and the political will to coordinate youth development programs and failed to specifically address gangs. As a result, efforts to deal with gang members under the age of eighteen through rehabilitation have fallen short as well and facilities for this purpose have facilitated “even deeper involvement in violence and crime.”

The failure of Mano Amiga policies in El Salvador across these areas comes down to a lack of resources and political will. In Honduras, the outlook of these policies is similarly grim. Partially as a result of international involvement, there existed a distinct lack of institutional coordination between individual programs. This lack of coordination resulted in inefficient initiative operating in isolation unable to learn from failed programs or strategies due to a lack of inter-organization communication. As a result of being underfunded and under-resourced in favor of the more politically popular gang suppression measures, the Mano Amiga policies have been limited in their effect.

In comparison to El Salvador and Honduras, Guatemala has more effectively and more widely implemented these types of policies. These prevention and intervention policies are credited by many as the reason that Guatemala has the lowest homicide rate in the Northern Triangle. Guatemala’s success illustrates that, if given the necessary political support and resources, the Mano Amiga policies provide a valuable avenue for improving security.

The failure of Mano Amiga policies in El Salvador across these areas comes down to a lack of resources and political will. In Honduras, the outlook of these policies is similarly grim. Partially as a result of international involvement, there existed a distinct lack of institutional coordination between individual programs. This lack of coordination resulted in inefficient initiative operating in isolation unable to learn from failed programs or strategies due to a lack of inter-organization communication. As a result of being underfunded and under-resourced in favor of the more politically popular gang suppression measures, the Mano Amiga policies have been limited in their effect.

In comparison to El Salvador and Honduras, Guatemala has more effectively and more widely implemented these types of policies. These prevention and intervention policies are credited by many as the reason that Guatemala has the lowest homicide rate in the Northern Triangle. Guatemala’s success illustrates that, if given the necessary political support and resources, the Mano Amiga policies provide a valuable avenue for improving security.

Figure 6: Data from 2017 LAPOL report on El Salvador

Figure 19. Percentage of respondents who report police engagement with the community, El Salvador 2014-2016

3.3 La Tregua

Following the lackluster results of the Mano Dura and Mano Amiga policies, both El Salvador and Honduras experimented with negotiating directly with gangs as an alternative approach. Gang leaders in Honduras attempted to replicate the effort in 2013, but the prime example of these negotiations is the 2012 gang truce in El Salvador. In this case, gang leaders from both MS 13 and Barrio 18 negotiated with the government to be
transferred to a less secure prison and were given $25 million in exchange for cooperating with the government to reduce the level of violence within the country. The truce yielded immediate benefits as the intentional homicide rate in El Salvador fell from 70.6 to 41.7 homicides per 100,000 people from 2012 to 2013, a reduction of over 40 percent.⁵⁷ Although El Salvador benefitted from a temporary lull in the homicide rate, when the truce collapsed in mid 2014, violence reached new highs and the homicide rate surpassed the 2012 rate.⁵⁸

While the number of cases is limited, the evidence is clear that cooperation and negotiation with gangs is not a sustainable policy. Despite the appeal of an immediate reduction in violence, both cases have illustrated that the truces only reduce violence in the short term and that this improvement likely comes at a significant long-term cost to the state and society. Gangs and politicians enter into these agreements when it is in their best interest; however, these interests of these actors do not necessarily align with what is best for society. By negotiating directly with criminal organizations, governments legitimize gangs’ political power and their status as societal actors while establishing the precedent that criminal gangs can gain political capital through the use and constraint of violence.⁵⁹ Both the 2012 truce in El Salvador and similar negotiations between the Brazilian government and the PCC appear to have strengthened gangs in the long run by improving group cohesion and giving them a reprieve from the pressure of prosecution.⁶⁰ Ultimately, these negotiations have demonstrated to the gangs that they can leverage their ability to inflict violence to produce political changes and manipulate state action to their benefit. Since the collapse of the 2012 truce in El Salvador, the gangs have become more aware of the political capital of their violence. This was illustrated by their 2016 attempts to leverage a brief ceasefire for political concessions.⁶¹ As gangs leverage their political capital more frequently, governments must take action to counter the influence and power of gangs because, unchecked, the gangs will assert political power to the detriment of states and societies across the Northern Triangle.⁶²

3.4 Colombia

Prior to the development and implementation of the DSDP, Colombia followed similar approaches towards the FARC and achieved results similar to those of the Northern Triangle countries. For the majority of the FARC insurgency, the Colombian government deployed the police and the military to directly attack the FARC and those who supported them.⁶³ Similar to the Mano Dura policies, these attacks in isolation did not lead to success and allowed the FARC to increase its strength and grow support through the 1990s.⁶⁴ The Colombian government also attempted to negotiate with the FARC at multiple points, most notably from 1992 to 2002 under President Pastrana. These negotiations included the creation of a demilitarized zone that the FARC took advantage of by using it to hold hostages and train troops for future offensive actions against the government.⁶⁵ These negotiations did not lead to peace and instead gave the FARC the time and space necessary to grow their strength.⁶⁶ The failure of this approach in Colombia set the stage for the implementation of a counterinsurgency campaign in 2002. These failures and the successful readjustment of policy towards a more comprehensive approach to security provide the framework for a policy comparison between Colombia and the Northern Triangle.

3.5 Conclusion

Despite efforts from the governments of the Northern Triangle, violence remains endemic and the underlying factors are under addressed. While the policy efforts have not been as successful as imagined, they have provided valuable lessons for informing future policy that should be implemented on a broader scale with greater intragovernmental coordination in order to have desired effect. The Mano Dura policies illustrated that suppression tactics alone are ineffective and can even be counterproductive if not used with precision and as one component of a broader approach. The Mano Amiga policies offer an alternative for addressing violence in more comprehensive ways but showed these types of policies must be properly resourced and supported by elected leaders. La Tregua, the attempts to negotiate with gangs showed that, while politically expeditious, these temporary reprieves strengthen gangs and make them a larger threat to both the state and society in the
long run by legitimizing them as political actors.

**Section 4: Outlining an Integrated Security Strategy**

The Northern Triangle should develop and implement an ISS that builds both on previous policies and the example provided by Colombia. Colombian COIN shows that an ISS should consist of four parallel lines of effort working together towards a common goal. The parallel lines of effort will marginalize violent actors and empower citizens. This section outlines these lines of effort, explains their importance and draws upon the Colombian case to give examples of what they would look like. Though examples from Colombia will not always be directly applicable, they offer jumping off points for future policy prescriptions.

Given the historical legacy of COIN in the Northern Triangle, the parallel drawn between an ISS and COIN merits clarification. While an ISS would be similar to COIN at the strategic level because it seeks to address the underlying factors that drive the security threat, an appropriate ISS for the Northern Triangle will have significant operational and tactical differences from the Colombian case because the gangs of the Northern Triangle do not employ the same type of violence as traditional insurgents. Recognizing this reality and abiding by the law of proportionality, the level of violence security forces should employ must therefore be different.

**4.1 Governance**

The first line of effort in an ISS for the Northern Triangle should be improvement of governmental coordination between institutions because this is a prerequisite for more effective policies. Government coordination is crucial because a whole of government response is needed to address the different economic, social and political sources of insecurity. As such, coordination will require some framework for information gathering, information sharing, and interagency coordination. This cooperation should be institutionalized at all levels to ensure input from appropriate stakeholders and to develop appropriate policies. The creation of structures for informing and executing policy is more important than any specific piece of substantive policy. This is because, as articulated in the problem driven iterative adaptation framework, the process of trying, learning, iterating, and adapting is crucial for addressing the complicated problems in play. In addition to government unity, government efficacy must be improved to actually carry out government policy. This would consist of investments to professionalize the civil service as well as anticorruption efforts to reduce waste and enhance legitimacy.

In the case of Colombia, significant effort was dedicated from the outset to improving both the coordination of government action by different government actors and the effectiveness of governmental institutions to implement new policies. To improve governmental coordination, Colombia’s DSDP created the National Defense and Security Council. This ministerial level forum coordinates the development and implementation of security policy across government agencies. New joint intelligence committees at the national and sub-national levels complemented the security council by coordinating the different intelligence and information collection capabilities of the state. These intelligence committees produce consolidated intelligence that better informs decision makers at all levels, thereby increasing the effectiveness of interdiction efforts by security forces. Colombia also took significant steps to improve governmental effectiveness by improving the “professionalization and specialization of state employees” through international cooperation and investments in education. Colombia also focused on improving the transparent efficient use of government resources. As discussed by Patrick Quirk and demonstrated by the World Bank’s index for the effectiveness of governance, these policies contributed to significant improvements in governance in Colombia.

**4.2 Security**

The establishment of state authority and the provision of security to all citizens is both a central objective of an ISS and a component of the ISS itself. As a component of the ISS, this initiative should include measures to suppress crime and marginalize gangs as actors by investigating criminal activity, prosecuting crime within the court of law and punishing those convicted of crimes. While the Northern Triangle countries have made improvements in this regard, the improvements have been insufficient. Improving security will entail improving the training and equipment for security
forces, expanding the investigative capacity for both criminal activity and improved investigation process for abuses of power and corruption. In addition to the security forces themselves, accomplishing these goals will require improvements to the judicial system so that cases can be adjudicated fairly and quickly. Moreover, prison reform will be needed to reduce overcrowding and limit the ability of gangs to use prisons to their advantage. Second, efforts should be made to prevent criminal activity wherever possible through community engagement and community intervention. In order to meet these objectives, major improvements will need to be made to the state’s security forces, investigative forces and court systems.

Figure 7: Data from World Bank World-Wide Governance Indicator Dataset

Colombia pursued the improvement of security and the expansion of the rule of law as a central component of the DSDP, including several steps intended to improve security through the marginalization of violent actors. While emphasizing the “respect, defense and promotion of human rights,” Colombia expanded the national police and centralized them within the comprehensive security entity of the Ministerio de Defensa. Although this did not completely eliminate ‘false positive’ killings, it did reduce the rate of human rights complaints and establish norms within the Ministry of Defense against human rights abuses that are evident today [Figure 8]. Within the expansion of security forces, Colombia dedicated specialized forces to pursuing the leadership of the FARC and the sources of their revenue. While these targeted attacks did not singlehandedly defeat the FARC, they reduced the cohesion, resources and organization of the FARC.

To complement expanded and improved security forces, Colombia also took significant steps to transition the justice system from an inquisitorial one to a accusatorial one while improving the investigative functions of the state. Because of improvements in efficiency and greater rights for defendants, the performance of the accusatory system has been “highly positive.” Complementing the more suppressive elements of the security campaign, the Colombian government also worked closely with communities to implement prevention and intervention programs. Colombia recognized that “prevention based on citizen participation is more effective than coercive measures to contain crime.” Previous attempts to do this had created autonomous self-defense groups and opened the door to well documented violations of human rights. Instead, the DSDP demobilized these groups so that the state could directly control all security forces. To complement state security forces, Colombia encouraged “the participation and co-operation of each and every citizen in the achievement and maintenance of security.” The DSDP did this by creating a neighborhood watch system and incentivized participation through a reward system for those who provided information that prevented attacks. As a result of the new approach to security in the DSDP, security has improved significantly in Colombia. The homicide rate has been cut in half while terrorist attacks and kidnappings have fallen by 90 percent.

Figure 8: Data from Colombian Ministry of Defense on human rights violations

4.3 Economic empowerment

The improvement of economic conditions and the creation of better opportunities for citizens of the Northern Triangle is a crucial line of effort within a broader ISS because new economic opportunities reduce the
allure of the illicit economy. New economic opportunities should be generated by improving business conditions, developing human capital and investing in economic infrastructure. Initial improvements in security, rule of law and governance are necessary to collectively improve business conditions and facilitate foreign direct investment. Given that crime-related costs reduced GDP by 5 percent in 2014, even incremental improvements to security and the rule of law would have significant economic benefits. The development of human capital through improvements to education can complement more general improvements to business conditions because it will increase economic opportunities. This is particularly important at the level of primary education, an institution that continues to be underfunded despite evidence showing it provides the best return on education investment and has been proven to reduce crime in the Northern Triangle. Finally, investments in infrastructure and public goods should be made to improve long-run economic growth.

Both Colombia’s DSDP and subsequent government actions at the national and sub-national level improved economic conditions and facilitated future growth. First, improvements to the rule of law and security dramatically improved economic growth. Some have even gone so far as to characterize these improvements as “the door through which much followed” because of Colombia’s economic gains from 2002 to 2014. In addition to structural conditions, Colombia made significant investments in infrastructure and economic development to build upon improvements in security conditions. An example of a successful initiative that accomplished these objectives is the Medellín Gondola which connected the less developed parts of the city with the vibrant center areas. While this infrastructure project was not tremendously flashy, it generated significant economic improvements while also increasing sociopolitical inclusion and improving security in a crucial part of the country. Beyond specific projects, Colombia also dedicated significant efforts and forces to protecting the economic infrastructure of the country from attack and these efforts helped encourage the investment of foreign capital in Colombia.

**Figure 9: Upper bound of crime-related costs as reported by UNODC**

[Graph showing crime-related costs as a percentage of GDP, 2014]

4.4 Sociopolitical Inclusion

Recognizing that previous policies have failed to address sociopolitical exclusion as a driver of gang membership, an ISS should specifically address sociopolitical exclusion. An ISS would do this both by improving participation from diverse segments of civil society
while also reintegrating gang members into society. Doing this would break down the social legitimacy gangs have tried to foster and decouple them from the communities they rely on. Diverse segments of society can be empowered by creating forums and structures for including these different stakeholders in the policymaking process. Expanding the participation in the process of developing and implementing policy is important because it illuminates the unintended effects of government policies while highlighting the interests of marginalized groups. While improving participation and inclusion, it is also important to protect the community leaders who participate in this process of consultation from retribution or threats from the gangs who have an interest in undermining this process. An ISS should also take steps to reduce the stigmatization of gang membership so that gang members can leave gangs and be reintegrated into society as gangs become a less attractive option. This reintegration of gang members is crucial because it is more sustainable and more efficient than prosecuting and incarcerating all gang members.

Recognizing the importance of sociopolitical inclusion in the broader effort to improve security, Colombia sought to “encourage the participation and co-operation of each and every citizen in the achievement and maintenance of security.” In addition to directly involving communities in security efforts through the creation of local self-defense groups, Colombia also included communities by creating regional security councils to oversee policy implementation in their respective areas. These councils and advisory committees were composed of “academics, businessmen and members of civil society” who advised policy makers, contributed “know-how” and represented voices that might not otherwise have been represented. Importantly, these mechanisms for participation were complemented by an expansion of the interior ministry’s program for persons at risk. By providing additional security for those participating in the policy making process, this change helped protect community leaders and political voices that might have otherwise been threatened.

Section 5: Conclusion

This piece examines the need for an ISS for the Northern Triangle and outlines what this ISS would entail. To do this, I analyzed the underlying sociological factors driving violence in the Northern Triangle as well as the role of third-generation gangs in limiting the government’s response to these factors. Based on this analysis, an assessment of previous policy approaches to these issues, and an examination of the Colombian case, I outlined a four-pronged ISS that improves security by marginalizing criminal actors, and addressing the underlying factors driving violence and criminality in the Northern Triangle. While the development and implementation of an ISS would not be an easy, quick or politically expedient process, it would be well worth the investment in the long run.

Salvadoran president Bukele’s Plan Control Territorial represents a significant shift towards the comprehensive approach to security envisioned in the concept of an ISS. Unlike Mano Dura, it acknowledges some of the drivers of violence and the importance of reasserting state control over territories occupied by gangs. This new approach is one of the reasons homicide rates have begun to fall in El Salvador from their peak. However, even this plan stops short of addressing the sociopolitical dynamics of security that will continue to complicate the situation if they not sufficiently addressed.

Acknowledgments

I would like to thank Dr. Cynthia McClintock, Dr. David Spencer, Emily Smith, Wazim Mowla, Jacob Mazzarella, Philipp Eibl and the team at the Columbia Journal of Politics and Society. Their contributions and support throughout research and writing process were immeasurable.

Endnotes

8. Ibid.
9. Ibid.
20. Barnes, “Criminal Politics.”
24. Ibid.
26. Ibid.
27. Barnes, “Criminal Politics.”
Beyond Mano Dura

33. Hume, “Mano Dura.”
40. Ibid.
42. Dudley, “How ‘Mano Dura’ is Strengthening Gangs.”
46. Seelke, “Gangs in Central America.”
47. Moreno, “Mano Dura, Mano Amiga and La Tregua.”
50. Seelke, “Gangs in Central America.”
58. Ibid.
64. Ibid.
66. Davis et al., A Great Perhaps?
86. Davis, et al., A Great Perhaps?
87. Ibid., pp. 90-94.

Bibliography


Legitimidad-y-confianza.pdf.


“Honduras: State response to gang-related violence and criminal activity, including the ef-


hold/2016/04/02/79222748-f5c2-11e5-958d-d038dac6e718_story.html.


