

Walk Like a Man: An Experimental Evaluation of Leader Experience as Costly Signal

Daniel McCormack and Joshua Blank*

April 14, 2015

Abstract

Building on the growing literature on leaders in international relations, in this project we address the question of how leader experiences shape voter evaluations of potential leaders' "toughness." Drawing from a theoretical literature on signaling, we utilize a survey experiment to examine the effect of both military experience and gender on respondent evaluations of leaders' type. We first ask how military experience and gender interact to condition voters' evaluations of a potential leader's willingness to stand up to international adversaries, as well as how this evaluation affects candidate support. We find that while serving in the military increases respondents' evaluations of female candidates' crisis bargaining resolve, military service erases female advantages in other issue areas. Female veterans are also perceived less favorably than their civilian counterparts.

*Working paper. Please do not cite. Comments welcome.

This paper is motivated by two puzzles. First, how do voters prospectively evaluate candidates' crisis bargaining resolve? Scholars of international relations believe that "toughness" in international crises is important to understanding foreign policy outcomes during a leader's tenure. Yet the public opinion literature suggests that voters are ill-equipped to make complex judgment calls about many issues. Given the inexperience of many presidential candidates in the realm of foreign policy, candidates' "toughness" may be a particularly difficult dimension for voters to judge.

Second, why are female candidates often presumed to be less competent than their male counterparts in the realm of foreign affairs? Anecdotally one might recall the scorn heaped on Sarah Palin's assertion that she could see Alaska from her house; more systematically, Lawless (2004) finds that by a 61-3 margin, voters believe men are better prepared to respond to a military crisis than are women. In this paper we explore whether this advantage is inherent – that is, grounded in voter evaluations of men versus women – or whether it might instead be accounted for by an understanding of differential life experiences. Specifically, we evaluate under what conditions military experience communicates information about a leader's competence. We explore an experimental survey that allows us to test the extent to which military experience serves as a credible signal of a candidate's foreign policy competence. In most cases, we find that former members of the military enjoy no significant advantage over their civilian counterparts.

Consider, however, the case of Frank Thompson, a member of the infantry and Union spy during the United States' civil war. Our results indicate that had Thompson run for office following the Civil War, this military experience would have proved a significant campaign advantage. Thompson's campaign would have been highly unusual in the 1860s, however, given that Thompson's real name was Sarah Emma Edmonds. Our experimental results indicate that, just

as Edmonds disguised herself as a man to serve in the Union army, military service allows contemporary female candidates to “disguise” themselves as men on dimensions of national security competence. To preview more fully the results below, we find that voters do not perceive a difference on national security competence between men who have served in the military and men who have not. However, women who have served in the military are perceived to be significantly more competent on national security issues than women who have not. Crucially, however, these women are not perceived as stronger on national security than either men who have served in the military *or* men who have not. While military service allows women to separate themselves from other women, voters perceive them as identical to even the least resolute male type of candidate.

This article makes contributions to research in the study of both international relations and public opinion. First, we demonstrate that actors use a candidate’s military service and gender when prospectively evaluating that candidate’s foreign policy behavior. Not only do these two dimensions help shape evaluations of a candidate’s general foreign policy acumen, they also forecast voters’ expectations of a candidate’s behavior in specific foreign policy crises. We link these expectations to signaling models in which individuals can reveal their type by engaging in costly behavior – here, military service – that separates “weak” from “strong” types. Beliefs over leaders’ resolve have been linked to a wide range of interstate phenomena, from war to bargaining over the creation of international institutions.

Our research design provides a microfoundational test of the growing literature that proposes a link between individuals’ life experiences and behavior in office. The results suggest that manipulable life experience may matter less to behavior in office – or at least is perceived to matter less – than intrinsic qualities. We locate the advantage enjoyed by male candidates on

national security issues in voter perceptions of male national security competence rather than men's actual experience on national security issues. To the extent that members of the military are able to mobilize support on the basis of their service, our results indicate that this support is likely translated through another mechanism.

Second, our results also contain implications for the literature on gender and issue ownership. Specifically, we corroborate findings that females are evaluated more highly on both education and health care competence. However, these advantages are erased for females who have served in the military. Military service causes voters to attribute traditionally masculine traits to female candidates across the board. While this effect benefits women on national security issues, it disadvantages them on issues for which non-military females are seen as more credible.

Literature review

An enormous literature spanning domestic politics and international relations proposes a complex causal web of relationships binding domestic political competition, qualities of individual leaders, and the outcomes of crisis bargaining. In this section, we argue that despite directly implying a selection of leaders for their expected adroitness in crisis bargaining situations, these authors have left the process by which leaders are prospectively evaluated unexplained. Drawing from the public opinion literature, we suggest that because international negotiations are a particularly information-poor environment for voters, they often use heuristics based on candidate life experience to make decisions over who to support.

The most prominent arguments linking domestic political institutions to international be-

havior focus on how leaders' expectations of domestic punishment shape their behavior in office. For example, Bueno De Mesquita et al. (1999) argue that leaders are beholden to a group of individuals – the “selectorate” – and must convince some subset of this group to support their tenure in office. Because democratic leaders are evaluated primarily on their ability to provide public as opposed to private goods, these leaders are likely to try very hard to win the wars they fight. In turn, because democratic institutions incentivize an “all out” war effort, both democratic states and their adversaries are encouraged to find alternative mechanisms of dispute settlements (Bueno De Mesquita et al. 1999). Counterintuitively, this same logic implies that democracies are more likely to be the targets of military aggression: because they are sensitive to the costs of war, democracies rarely start fights (Filson and Werner 2004). ? find yet another mechanism for democratic peaceability: because democratic leaders are unlikely to be imprisoned or murdered when they leave office, they are less prone to undertaking high-risk strategies like interstate war in order to stay in office.

Democratic political institutions can also aid leaders against their international adversaries by supporting the leader's threat of war (Schultz 1998). In these cases, if an adversary is unsure over a leader's resolve to carry through with his or her threat, broad-based public support can operate to reduce this uncertainty. “Resolve,” in the context of interstate bargaining, generally refers to the cost a leader or country pays for going to war – if a leader's cost of war is low, (s)he is said to be highly resolved and willing to fight over a greater range of issues (and vice versa for leaders with high costs of war). Because a leader's resolve helps to determine what they are willing to fight over, uncertainty over this quality can generate a risk of war where otherwise none would exist (Fearon 1995).

The issue of international uncertainty over a leader's resolve opens the door to a wide range

of mechanisms through which domestic politics can shape international bargaining. For example, by making a demand of its adversary in view of its domestic public, a leader can tie his or her hands by making failing to follow through on the demand costly (Fearon 1994, 1997). When this occurs the leader can effectively promise to go to war even when (s)he would otherwise be unresolved.

One critical precondition for the existence of audience cost-type arguments, in which a leader uses his or her domestic public to display resolve to an interstate adversary, is that domestic audiences care about securing foreign policy goals. As illustrated above, most arguments linking domestic audiences to the outcome of international negotiations focus on ex post leader *punishment* strategies: if a leader fails to prosecute state goals abroad, (s)he can count on being removed or otherwise punished. But while they are relatively neglected in the international relations literature, ex ante strategies of leader *selection* should be just as effective as a way for domestic audiences to bind leaders to a certain policy. Here, rather than threatening to punish a leader should (s)he fail abroad, domestic groups might only elect leaders who they know are highly competent in foreign affairs.

In order for domestic audiences to select leaders based on their competence in foreign affairs, these audiences must have some sense of what makes an effective leader abroad. By and large this requirement is not met. Problematically, while voters in the U.S. appear to primarily evaluate candidates based on personal characteristics (rather than policy positions), the characteristics that draw the most attention are not those related to competence but rather personal character and physical attractiveness (Glass 1985). On the other hand, McDermott (2005) finds that voters are able to use candidate occupations to make guesses about their competence. In the realm of foreign affairs, there is not an obvious analogue to “foreign policy experience.” So

how do voters evaluate candidates' expected foreign policy performance?

The question of how previous life experience shapes a leader's behavior in office is a growing field. Recently, Horowitz and Stam (2012) find that the leaders most likely to engage in conflict while in office are those with military – but not combat – experience. Older leaders appear to be particularly likely to be targeted in militarized disputes (Bak and Palmer 2010). But do voters take these effects into account? At least one relationship is clear: voters give male candidates a heavy and consistent advantage when it comes to foreign policy evaluations. Lawless (2004) finds that respondents to a survey give men a 61-3 edge on the question of what type of candidate is best suited to respond to a military crisis. Part of this advantage may be due to an inconsistency between the female “sex role” and perceptions of how leaders are supposed to behave (Johnson et al. 2008). Integrating the discussion from earlier, voters may also perceive men to disproportionately have life experiences that conduce to positive foreign policy performance. If, as John McCain suggested, military experience prepares a candidate for office, perhaps women are simply perceived in the aggregate as insufficiently prepared for political leadership due to their experience deficit. In the following section we outline a test to examine this proposition.

Theory

The quintessential problem facing voters is how to determine a candidate's worth in an information environment where candidates have an incentive to be intentionally misleading. Because voters prefer a competent leader to an incompetent one, all types of leaders are willing to state that they are competent; consequently, voters cannot believe them. In this section,

through the description of an informal theoretical model, we provide a signaling explanation for why military service may reveal to voters a candidate's foreign policy competence.

Suppose a voter has a choice between two candidates, a "tough" and a "soft" candidate. The two types – soft and tough – are distinguished by how willing each is to engage in conflict. Borrowing from the formal literature on interstate conflict bargaining, the "tough" type pays a small cost for going to war, while the "soft" type pays a larger one. When these costs are fully known, being more willing to go to war does not translate into a higher probability of engaging in war. Rather, because a potential leader's adversaries observe his or her level of resolve, a "tough" leader can secure better bargaining outcomes for his or her country.¹ Under these conditions, a voter will always choose the tough type of candidate.

The foregoing underscores the key logic of a potential leader's dilemma. Because voters will choose leaders they believe to be tough, a candidate has an incentive to portray him or herself as tough, even when (s)he is not. Anticipating this selection process, a potential leader might seek to credibly convey to voters his or her toughness. Paraphrasing from Fearon (1997), "For a [signal] to increase the [voter's] belief that the [candidate] would be willing to fight, it must be more likely that a resolved [candidate] would make the [signal] than an unresolved [candidate]. Thus, to be credible, a [signal] must have some cost or risk attached to it that might discourage an unresolved [candidate] from making it" (p. 69).

Building from this logic of costly signaling, we propose that voters perceive military service to operate as a credible signal of a candidate's toughness. That is to say, voters perceive a leader that has served in the military to be "tougher" than one that has not. In the following section, we

¹What we are *not* interested in here is the possibility of preference divergences between a leader and his or her selectors.

design and describe an experimental survey design that allows us to test precisely this proposition.

Research Design

The following experiment was conducted in the Fall of 2014 on 600 undergraduate students at the University of Texas at Austin taking an introductory course in government. The experiment was administered using Qualtrics survey software. The students were provided with brief instructions and a link to the survey, at which point they were given a week to complete the assignment for nominal extra credit.

Conducting the experiment in this fashion has its share of advantages as well as disadvantages. The advantage of conducting an experiment online and at a time of the subject's choosing is found in the reduction in infrastructure costs associated with a laboratory experiment in addition to the ability to increase one's sample size beyond the structural limitations of a lab. The major disadvantage is the loss of environmental control. In some instances, this is rather innocuous. Some subjects will face more distractions than others – a real life consideration – depending on their chosen location for completion, but this potential problem should be randomly distributed across the treatment and control conditions. In addition, the nature of the experimental manipulation (to be described below) is that it relies on the processing of manipulated information found in a fictitious news article, not the arousal of some underlying cognitive or emotional state. If we were seeking to do the latter, the distractions associated with a loss of environmental control would be far more damaging. More problematic than the potential for distraction is the fact that subjects might begin the experiment, stop for an infinite

number of reasons and for varying amounts of time, and then resume, weakening potential treatment effects; or, simply click through the experiment as quickly as possible, negating any treatment effects.

To combat this possibility, the time elapsed to complete the experiment was recorded for each subject (as well as a number of inter-experimental times on treatment and control durations). Given that the distribution of the time elapsed for subjects was approximately normal, we decided to remove subjects from the analysis if their time to completion was greater than 1.5 standard deviations from the mean time to completion. This resulted in the exclusion of 105 students.²

For the remaining 600 students, the experimental manipulation consisted of introducing subjects to a fictitious candidate for the 2016 Presidential Election through a short news article crafted to resemble a piece on the website Politico. Subjects were exposed either to a candidate who had served in the military or one who had not, and to a candidate who was either male or female, making this a 2x2 factorial design.

The candidate (Allen or Alice Smith), a member of the House of Representatives from Michigan and the former mayor of Lansing, is noted in the article because of increasing chatter about their candidacy, but also because of tough comments made towards Vladimir Putin during a committee hearing on sanctions for Russia's recent actions.³

Because party identification is such an important driver of attitudes, we first asked subjects about their party identification and then filtered the news article such that Rep. Smith would be presented as a member of the subject's stated party (Campbell 1960).⁴

²In concrete terms, 34 students spent less than 28.5 seconds on the entire experiment, while 71 students spent more than 343.5 seconds.

³The full text of the article and the manipulations can be viewed in the appendix.

⁴Subjects who associated with neither party were randomly distributed between the treatment. This should

After reading the article, subjects were asked a battery of questions to evaluate the candidate, and in particular, the candidate's toughness. Subjects were asked, "*If elected president, how effective do you think Representative Smith would be in standing up to Vladimir Putin?*" and were given the response options "*Extremely effective*", "*Somewhat effective*", "*Not very effective*", or "*Not at all effective*". In addition to Putin, subjects were asked the same question with regards to Iran and ISIS. The timing of the survey experiment – it was administered the week of October 20, 2014 – was fortuitous. This week saw both the advance of ISIS to within several dozen miles of Baghdad, as well as the hotly contested run-up to the Ukrainian parliamentary elections (to which concerns over Russia's involvement were intimately connected). Finally, this period also overlapped with intensified international negotiations over Iran's nuclear program in advance of a self-imposed deadline of late November. In sum, all three potential adversaries were likely highly salient in respondents' minds.

The purpose of the toughness battery was to assess whether perceptions of the candidate's toughness changed based on their gender and their background. In choosing these three objects (Putin, Iran, and ISIS), the intention was multi-fold. First, we are able to directly assess whether Rep. Smith's tough words toward Putin are reflected in the subjects' evaluations given Smith's gender and background, but also whether these tough words were generalized to the other objects for which the Representative made no mention. In addition, our intention was to choose objects with some conceptual variance between them: Putin, an individual; Iran, a state; and ISIS, a non-state organization. Responses were coded from 0 to 1 so that higher values indicate greater perceived effectiveness.

randomly distribute this group among the conditions. It should also be noted that differences in results by partisanship, though interesting, are not a primary focus of this research.

In addition to this primary battery, subjects were also asked to evaluate the candidate on more general criteria. To this end, subjects were asked, “*In general, how effective do you think Rep. Smith would be at handling the following issues if elected president?*”, and then asked to evaluate Smith on the same scale as above, but for the issues of national security, education, healthcare, and the economy. National security is the issue most relevant to subjects’ assessments of candidate toughness, but in addition, we wanted to assess whether altering gender or background lead to different assessments of the candidate across a range of policy areas to see whether the gender/background effects were focused or diffuse. Again, responses were coded from 0 to 1 so that higher values indicated greater perceived effectiveness.

Finally, subjects were asked their overall favorability towards the candidate. Subjects were able to indicate whether they viewed the candidate *very* favorably, *somewhat* favorably, neither favorable nor unfavorably, *somewhat* unfavorably, or *very* unfavorably. Responses were coded from 0 to 1 so that a higher score indicated a more favorable view.⁵

One final element of the experimental design was to test whether audience costs are distributed differently based on gender and background. After answering the questions described above, subjects were presented with a second, short Politico article in which Rep. Smith retracts his/her tough language towards Russia (see appendix).

After reading this article, subjects were again asked to indicate their attitude toward the Representative on a standard favorability item as described above. Differences between the

⁵It should be noted that the order of the questions as presented here was chosen for conceptual clarity and research importance. With respect to their actual ordering, subjects were first presented with the basic favorability item, then the toughness battery, and finally the broad issue areas. The logic for this ordering is built on best practices in survey design in conjunction with a decision about our primary variables of interest. The favorability item is general in nature, making the results easy to modify if placed after the toughness questions based on the logic that the toughness questions would then become the criteria by which subjects would choose to evaluate the candidate on the favorability item. The more specific toughness battery came prior to the more general issue battery because of its centrality to our research question.

initial favorability item and the post-retraction favorability item are calculated and used as an indicator of audience costs. We now turn to the results.

Results

The results section is divided into three main sections. In the first, we highlight the main findings that link candidates' gender and military background on to voter evaluations of their toughness. In general, these results show two patterns. As expected, voters perceive female candidates to be less resolved against all foreign policy adversaries than male candidates. While military service tends to increase voter evaluations of candidate resolve, the interaction with gender is more nuanced. Across all three adversaries – Putin, Iran, and ISIS – women with combat experience receive a statistically significant bump in evaluations of their resolve over their civilian counterparts. While male veterans appear to be perceived as more resolved than male non-veterans, this difference is not significant. Intriguingly, voters perceive female veterans to have statistically indistinguishable levels of resolve relative to either type of male candidate. In short, military service allows a female candidate to “walk like a man” in terms of her foreign policy credibility.

The second section turns to an evaluation of the treatment on voter evaluation of candidate issue competency. Here we find that for national security, the interaction of gender and military service has much the same effect as before. Female veterans appear similar to both types of male candidates, but more credible than their non-veteran female equivalents. We also show that military service effectively *erases* perceived female competency advantages in education and healthcare. Just as military service allows women to look like men (to their advantage) in

the realm of national security, it has the same effect to their *disadvantage* in issue areas where female candidates traditionally have an edge.

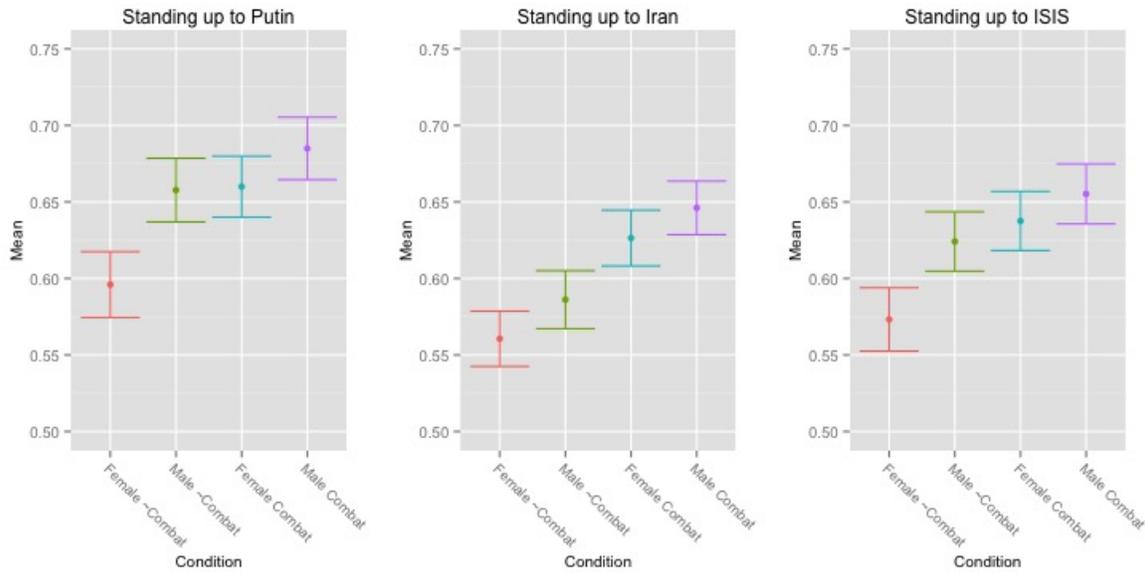
Finally, we provide a look at how gender and military service helps to shape overall candidate favorability. Troublingly, we find that female veterans are perceived much less favorably than either female civilians or male veterans. The net effect of a boost in perceived national security competency and a hit to education and healthcare competency is to dramatically decrease voter evaluations of female candidate favorability.

The presentation of these results takes a persistent format. We examine results for each item with a one-way analysis of variance (ANOVA) to test whether differences exist in the data generating process between the four treatment groups. If we find differences, we then compare means and perform a post-hoc pairwise *t*-test to see which pairs present the statistically significant differences.

We begin first with the toughness battery. A one-way ANOVA indicated a statistically significant effect of treatment condition on effectiveness in standing up to Putin ($F(3, 572) = 3.18, p < .05$), Iran ($F(3, 572) = 4.38, p < .01$), and ISIS ($F(3, 572) = 3.04, p < .05$). Figure 1 displays the results for each of the tough type battery objects. As expected, the cleanest break between female non-veterans and the other candidate types appears in the battery relating candidate toughness to negotiations with Vladimir Putin, the subject of the Politico article. However, the general pattern – perceived weakness for female civilians and statistical indistinguishability for the remaining types – persists across all three adversaries.

When it came to assessing the Representative's effectiveness in standing up to Putin, the female candidate without combat experience was rated as less effective than all other potential candidates: the male without combat experience ($p = .034$), the female with combat experience

Figure 1: Tough Type Battery



($p = .031$), and the male with combat experience ($p = .003$). Neither the male nor the female with combat experience were perceived as significantly more effective than the male without combat experience.

In the Iranian case, assessments of the effectiveness of the male and female candidates without combat experience were statistically indistinguishable, but the female with combat experience was assessed as likely to be more effective than the female without combat experience ($p = .012$), and her results approach statistical significance when compared to the male without combat experience ($p = .11$). The male with combat experience was deemed to be more effective than the female without combat experience ($p = .001$) and the male without combat experience ($p = .019$), but not the female with combat experience ($p = .439$).

A similar pattern held true for ISIS. The female candidate without combat experience was perceived to be less effective than the male without combat experience ($p = .072$), the female with combat experience ($p = .023$), and the male with combat experience ($p = .004$). The other three conditions were statistically indistinguishable. Overall, the three different potential for-

eign policy adversaries confirm the pattern described at the beginning of this section. Female candidates without combat experience are viewed as being unable to stand up to foreign policy adversaries. All other types of candidates – females with military experience and both types of males – are seen as roughly equivalent.⁶

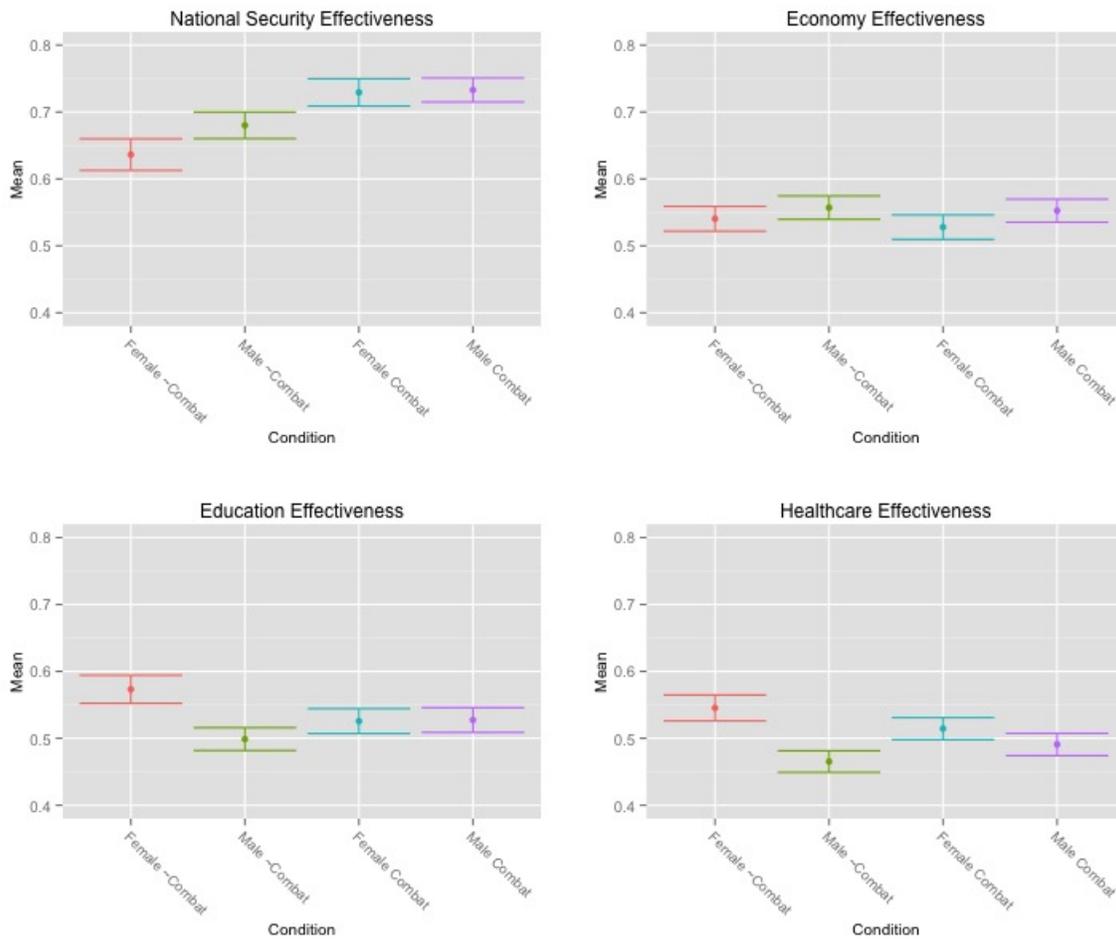
The next battery of interest is the issue battery. As a reminder, respondents were asked to rate the candidate's likely effectiveness as president on the issues of national security, the economy, education, and healthcare. We ran this battery of questions principally as a concept check: if voters in fact care about how "tough" candidates are, patterns of national security effectiveness should be broadly similar to those identified in the type tests presented above.

A one-way ANOVA test indicated a statistically significant effect of treatment condition on effectiveness in national security ($F(3, 572) = 4.84, p < .01$), healthcare ($F(3, 572) = 3.90, p < .01$), and education ($F(3, 572) = 2.61, p = .05$), but not for the economy.

A visual inspection of Figure 2 shows that perceptions of candidates' effectiveness in each condition was higher for national security than for any of the other issue areas. Overall, the effect of both gender and military experience on perceived national security effectiveness tracks well with the pattern we identified in the toughness battery above: females without combat experience are perceived as significantly less effective, while the remaining three types are statistically indistinguishable. The one exception to this trend is that the male and female candidates without military experience were statistically indistinguishable from each other. While the female with combat experience was perceived to be more effective than the female without combat experience ($p = .002$), she was not deemed more likely to be effective than the male

⁶The partial exception to this trend is for the male non-veterans in the Iranian case, in which they were viewed as equivalent to female non-veterans and female veterans and significantly weaker than male veterans.

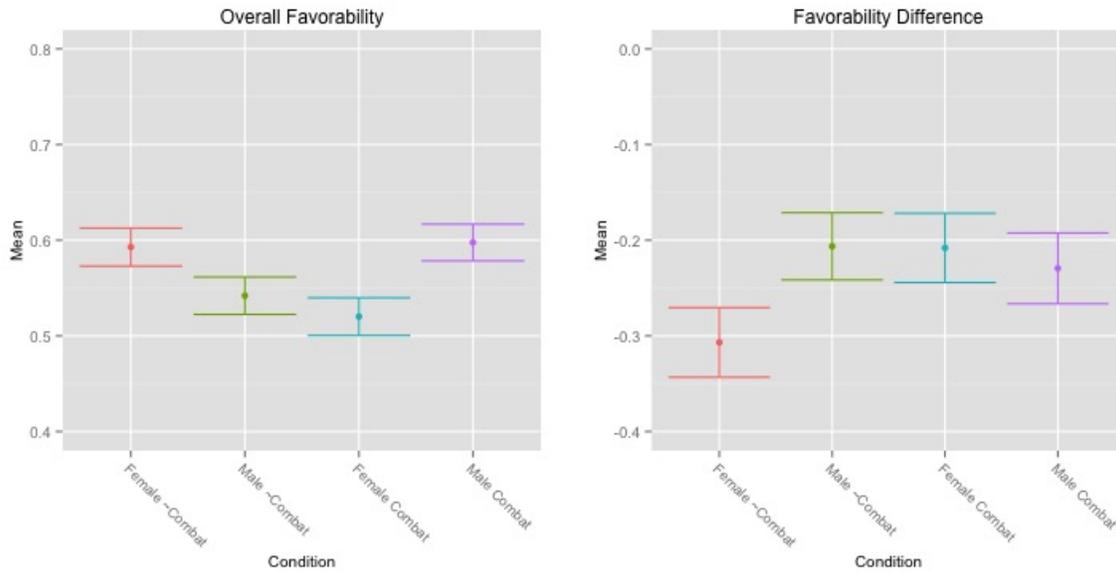
Figure 2: Issue Battery



without combat experience at conventional levels of statistical significance ($p = .083$). The male with combat experience was thought likely to be more effective than the female without combat experience ($p = .007$), and approaches conventional levels of statistical significance when compared to the male without combat experience ($p = .065$).

The general pattern that appears to govern national security falls apart entirely for the remaining three issue areas. As noted, there were no significant differences for any pairs for prospective evaluations of economic effectiveness. For both education and healthcare, the female without combat experience was rated the most highly of the four possible candidates. In education, this difference was significantly higher than a male without combat experience

Figure 3: Candidate Favorability



($p = .006$). The non-veteran female also enjoyed a positive, though not quite significant, differential over the female with combat experience ($p = .077$) and the male with combat experience ($p = .090$). On healthcare, the female without combat experience was expected to be more effective than both types of males, but indistinguishable from the female with combat experience ($p = .207$). In addition, the female with combat experience was expected to be more effective than the male without combat experience ($p = .038$).

The final question we were able to ask of the data pertains to the overall favorability of the candidate, both before and after the retraction of the strong statement against Vladimir Putin. Through a one way ANOVA of the favorability item by treatment group, we identified a statistically significant effect of the treatment on pre-retraction favorability ($F(3, 572) = 3.81, p < .05$). Interestingly, the female without combat experience and the male with combat experience were statistically indistinguishable, but were both rated more highly than the male without combat experience ($p = .071; p = .042$) and the female with combat experience ($p = .010; p = .005$). The male without combat experience and the female with combat experience were statistically in-

distinguishable, although the female with combat experience had the lowest mean favorability in the sample. The effect of serving in the military presents a stark difference by gender. While men receive a statistically significant boost from serving in the military, women receive a similarly significant decrease in favorability.

In order to calculate the favorability differences, we subtracted the pre-retraction favorability from post-retraction favorability. To be clear, larger values of this difference indicate that respondents punished the candidate to a greater extent for retracting his or her statement. The results of the one way ANOVA testing treatment group effect on these differences did not find a statistically significant effect by treatment group, although each type of candidate was punished by respondents for retracing their statement. A visual inspection of the right panel of Figure 3 shows that to the extent each candidate was punished, the female candidate without combat experience appears to have been punished more for retraction than the other candidates. This effect approached statistical significance when the non-combat female was compared with the male without combat experience ($p = .054$), the female with combat experience ($p = .058$), but not the male with combat experience ($p = .139$). Yet again, the overall impression from the data is that serving in the military allows female candidates to appear to subjects as though they were male.

Discussion & Conclusion

In this section we highlight two main findings by way of concluding. First, the effect of military service is to “mask” candidates’ gender. In general, veteran status does not boost respondent evaluations of female resolve past that of men. Theoretically, this result is surprising. A Bayesian

approach to signaling would expect that if voters believe female candidates to be weaker on national security – as they clearly do – then incorporating a signal “against their type” should lead to a larger treatment effect than it should for men. In other words, taking on a trait like military service should lead to a larger jump in evaluations of female resolve than male resolve, precisely because voters do not expect female candidates to take on the trait. In fact, if voters initially perceive female candidates to be less resolved than their male counterparts, military service should serve to boost them *ahead*.

The Bayesian story sketched above is resoundingly rejected by the results presented here. These results indicate that voter evaluations of candidate traits are likely driven by a perception of inherent, rather than manipulable, candidate traits. Rather than pushing female candidates ahead of male candidates in the realm of “toughness” or foreign policy competence, military service operates only to put female candidates on equal footing with their male counterparts. Taken alone, this effect might serve as indicative of the tentatively positive effect of military service. Yet taken in conjunction with the second primary finding, the effect of military service is considerably more troubling.

The interaction between the treatment effect on evaluations of candidate competency and evaluations of candidate favorability is complicated. While military service again operates to construct a respondent profile of female competence that mirrors that of a male, female candidates are made to pay in the realm of favorability. At least as it pertains to military service, female candidates can have either issue competency or favorability, but not both.

Moreover, not only are female candidates penalized in voter evaluations of their favorability due to their military service, men are *rewarded* for serving in the military. In sum, relative to their male civilian counterparts, male candidates are perceived to be weakly more competent

in all issue areas; more resolved vis-a-vis their adversaries; and more favorable. For women, veterans are perceived as more resolved and more competent in the realm of national security, but less competent in all other issue areas and significantly less favorable. While more research – and ideally a non-student sample – is necessary to continue to tease out the intricacies of voter evaluations of candidates’ competency and resolve, the results presented here point to one way forward to disentangling the effects of inherent and manipulable candidate traits, as well as their complex interaction with the role of gender in prospectively evaluating candidate behavior.

References

- Bak, Daehee and Glenn Palmer. 2010. "Testing the Biden Hypotheses: Leader tenure, age, and international conflict." *Foreign Policy Analysis* 6(3):257–273.
- Bueno De Mesquita, Bruce, James D Morrow, Randolph M Siverson and Alastair Smith. 1999. "An institutional explanation of the democratic peace." *American Political Science Review* pp. 791–807.
- Campbell, Angus. 1960. *The american voter*. University of Chicago Press.
- Fearon, James D. 1994. "Domestic political audiences and the escalation of international disputes." *American Political Science Review* 88(03):577–592.
- Fearon, James D. 1995. "Rationalist explanations for war." *International organization* 49(03):379–414.
- Fearon, James D. 1997. "Signaling Foreign Policy Interests Tying Hands versus Sinking Costs." *Journal of Conflict Resolution* 41(1):68–90.
- Filson, Darren and Suzanne Werner. 2004. "Bargaining and fighting: The impact of regime type on war onset, duration, and outcomes." *American Journal of Political Science* 48(2):296–313.
- Glass, David P. 1985. "Evaluating presidential candidates: Who focuses on their personal attributes?" *Public Opinion Quarterly* 49(4):517–534.
- Horowitz, Michael C and Allan C Stam. 2012. "How prior military experience influences the future militarized behavior of leaders." *International Organization* .

- Johnson, Stefanie K, Susan Elaine Murphy, Selamawit Zewdie and Rebecca J Reichard. 2008. "The strong, sensitive type: Effects of gender stereotypes and leadership prototypes on the evaluation of male and female leaders." *Organizational Behavior and Human Decision Processes* 106(1):39–60.
- Lawless, Jennifer L. 2004. "Women, war, and winning elections: Gender stereotyping in the post-September 11th era." *Political Research Quarterly* 57(3):479–490.
- McDermott, Monika L. 2005. "Candidate occupations and voter information shortcuts." *Journal of Politics* 67(1):201–219.
- Schultz, Kenneth A. 1998. "Domestic opposition and signaling in international crises." *American Political Science Review* pp. 829–844.