When human rights pressure is counterproductive: A survey experiment on women's rights in China

11,874 words

Jamie Gruffydd-Jones
Princeton University
22nd May 2016

In this article I examine when naming and shaming has counterproductive ‘backfire’ effects on public grievances about human rights. Using a survey experiment of Chinese netizens at the start of 2016, I show that when criticism of women’s rights in China comes from a threatening geopolitical opponent, especially when public nationalism is high, public grievances about women’s rights and willingness to petition for improvements in those rights both fall. Criticism that comes from a neutral source however does not see the same negative impact. These effects are dependent on the link between the government leaders and the nation however, and can be nullified and even reversed when the criticism specifically targets elites. The findings challenge one of the assumptions underlying models of human rights advocacy; that international pressure will benefit domestic efforts for change; and call for closer attention to the responses of citizens of target countries.

What is the public impact of efforts to shame regimes on their human rights? How do citizens respond to international pressure on their country? Most accounts of international human rights pressure fail to account for how policies will impact the domestic public of the target country. Studies on the widely established policy of naming and shaming for example examine, cross-nationally, the influence of the amount of shaming given to a country in one year on its human rights standards the following year, normally measured by changes on a 1-5 scale. These short-term behavioural measures are undoubtedly important, but their use risks overestimating short-term strategic concessions by under-pressure governments and ignoring potentially longer lasting public attitude changes. Given that the level of popular support for human rights issues in many cases has a significant impact on government policies, and that public grievances are actively built into many models of the impacts of the international human rights system, this seems a prominent oversight. Even in authoritarian countries such as China, government policy on issues like the death penalty has relaxed markedly in response to changes in public attitudes - in this case growing anger at high-profile wrongful executions.

In this article I argue that the public impact of information about international pressure on a country’s human rights conditions will depend on the perceived hostility of the source of the pressure to the target nation. Critical information about human rights that comes from a ‘benign’ source may have the positive effect on public grievances that

---

2 As noted by Hafner–Burton, E.M. (2008)
naming and shaming theorists expect. However when members of the public hear condemnation from a source that they believe has nefarious intentions, for example to use human rights to achieve some kind of geopolitical advantage, they will react defensively, strengthening their belief that human rights are good enough in their country. As a result I hypothesise that international pressure that comes from a threatening geopolitical opponent, especially when public nationalism is high, will have counterproductive effects on public grievances. Criticism at peaceful times from a neutral source, or that explicitly targets the elites of the country rather than the nation as a whole, will be much more effective.

I test my argument using a survey experiment representative of the online population on women’s rights in China in early 2016, immediately following the high profile closure of a centre for women’s rights in Beijing, which garnered domestic and international attention. The experiment randomly exposed respondents to criticism of women’s rights in China from either the United States, a geopolitical opponent, or the African Union, a neutral organisation and nominal ‘third world’ ally to Beijing. The experiment also manipulated the salience of the national identity in the respondents, by randomly exposing them to a Chinese flag, before then measuring their grievances towards women’s rights in China.

I find that, as expected, criticism from the United States significantly reduces Chinese citizens’ grievances about women’s rights conditions in their country. US criticism makes respondents 28.6% more likely to believe that women’s rights are well respected, as well as less willing to sign petitions to improve women’s rights. Criticism from the African Union on the other hand sees no effects in either direction. I also show that these effects can be reversed when the criticism targets the Chinese government specifically, rather than the country as a whole - suggesting that the bind between the elites and the nation is central to the counterproductive effect of criticism.

These findings have important implications for how we think about the impacts of international pressure on human rights in authoritarian regimes. Firstly, they challenge a central assumption of the literature on transnational activism - that pressure from foreign actors will dovetail with pressure from domestic actors to force changes in government behaviour. The study suggests instead that such a smooth interaction may be overambitious - and that when it appears to be threatening, foreign pressure may in fact make domestic pressure more difficult. Secondly, the findings redirect attention to how a target country’s public respond to international efforts to change behaviour on human rights. If foreign countries and organisations find success in securing prisoner releases and other short-term concessions, these are of little use without meaningful long-term policy changes, which depend strongly on societal pressures. Finally the results demonstrate the need to look more carefully at who exerts human rights pressure. If we care about public responses, geopolitical relationships matter, and pressure from the United States on its allies is likely to be much less successful than pressure on its competitors.

Public responses to international pressure

The literature on how publics respond to international pressure is only weakly developed. The only authors to address this issue, Davis et al. (2012) and Ausderan (2014), have shown that the amount of shaming of a country’s rights from groups like Amnesty International in one year has mildly negative effects on public perceptions of the standard

---

of human rights conditions the following year\textsuperscript{7}. They argue that when citizens receive information about shaming of rights abuses in their country, they update their perceptions of how much their country respects human rights - or at least when there are NGOs in-country to help circulate that information\textsuperscript{8}.

This supports the assumption that underlies many of the models of human rights pressure; that information about international pressure will have a positive effect on domestic movements for change\textsuperscript{9}. In the models shaming influences target regimes directly, but also indirectly, providing information about abuses and foreign condemnation to the domestic public and civil society. According to Davis, Murdie and Steinmetz “dissemination of negative information about a government, is an integral part of the process by which individuals find out about government repression”\textsuperscript{10}. Increased public knowledge of government repression and foreign opprobrium will help drive support for domestic opposition groups. At the very least, a central tenet of the two most comprehensive stories of human rights change, the Boomerang and Spiral models, is that pressure from above and pressure from below work in tandem, foreign and domestic actors providing assistance to each other in challenging the government\textsuperscript{11}. For Risse, Ropp and Sikkink, this is the only way that we can expect improvements in human rights\textsuperscript{12}.

On the other hand, others have suggested that the diplomacy of shaming may arouse anger rather than shame\textsuperscript{13}, and trigger a defensive nationalist backlash\textsuperscript{14}. Studies that have taken a more nuanced look at the public response to foreign endorsements have used experimental surveys to uncover the impact of different kinds of intervention. These have suggested that as far as support for electoral candidates\textsuperscript{15} and support for women’s participation in politics\textsuperscript{16} go, foreign intervention tends to have a polarising effect, depending on degree of existing support for the regime; - and that where the information comes from seems to make very little difference.

In this study I focus on ‘shaming’ - the criticism of human rights conditions - rather than ‘naming’ - the detailing of specific abuses themselves. As ‘naming and shaming’ suggests, the naming of abuses is invariably accompanied by criticism, and so while the theory I provide should also extend to revelations of specific abuse (as provided by Amnesty International or media reports for example), it should be noted that information about specific violations that might accompany this criticism would most likely lead to different responses than shaming alone. Much of the information that is reported in news items about international pressure on human rights is likely to focus on the shaming aspect however- news about criticism of conditions and calls for improvement. This kind of ‘commentary’ rather than ‘revelatory’ information is also the most likely to come from states or international organisations. In Spring 2016 for example, in a number of countries

\textsuperscript{7} Ausderan, J. (2014); Davis, D. R., Murdie, A., & Steinmetz, C. G. (2012)
\textsuperscript{8} Davis, D. R., Murdie, A., & Steinmetz, C. G. (2012)
\textsuperscript{10} Davis, D. R., Murdie, A., & Steinmetz, C. G. (2012) pg202
\textsuperscript{12} Risse, T, Ropp, S.C., & Sikkink, K (1999), pg276
\textsuperscript{13} Wachman, A. M. (2001). Does the diplomacy of shame promote human rights in China?. Third World Quarterly, 22(2), 257-281
\textsuperscript{14} Carothers, T. (2006) The backlash against democracy promotion. Foreign Affairs, 85(2), 55-68

There are also design reasons for this choice. In an experiment of this sort, providing generic information about an instance of foreign criticism allows more generalisability about responses to criticism. Specific information naturally differs from case to case in its evidential value and novelty, making it hard to extend beyond the information used.

**National Identity Threat and Public Responses to Naming and Shaming**

Scholars of attitude change have noted the existence of a ‘boomerang’ or ‘backfire’ effect\footnote{As not to confuse with Keck and Sikkink’s ‘Boomerang model’, I use the ‘backfire’ terminology}. This is when a message that is designed to change attitudes or behaviours in a certain direction ends up pushing those attitudes or behaviours in the opposite direction; a phenomenon noted in areas from anti-litter campaigns\footnote{Reich, J. W., & Robertson, J. L. (1979). Reactance and Norm Appeal in Anti-Littering Messages. \textit{Journal of Applied Social Psychology}, 9(1), 91-101.} to climate change. Nyhan & Reiffler show that when conservative Americans were presented with evidence showing that President George W. Bush’s statements about Weapons of Mass Destruction (WMDs) in 2004 were in fact false, they became even more confident in their misperceptions, ending up more likely to believe that Saddam Hussein was in fact stockpiling WMDs.\footnote{Nyhan, B., & Reiffler, J. (2010) When corrections fail: The persistence of political misperceptions. \textit{Political Behavior}, 32(2), 303-330.} Nyhan and Reiffler argue that this may be because, in seeing information that goes against their strongly-held beliefs and political identities, conservative Americans then spend time generating arguments that might counter this information, and as a consequence strengthen their beliefs.\footnote{Following Byrne and Hart (2009)}

Just as conservative Americans strengthened their beliefs that Hussein was stockpiling WMDs following information to the contrary, I argue that following critical information about human rights violations in their country, citizens with some level of national attachment will (under certain circumstances) strengthen their beliefs that their country’s human rights are good enough. I argue that this is because when hearing information that is critical of their national identity, people do not only reject the information, but are motivated by the threat to their identity to create counter-arguments, searching their memory for evidence that their country does not violate human rights. As long as there is some information available to be found, this creates a backfire effect and grievances about human rights go down.

that threatens their identity in some way - for example that Republican leaders made an error in going to war over WMDs - then they will be motivated to defend themselves and resist this new information. As Sherman & Cohen argue: “people will defend against threats to collective aspects of the self much as they defend against threats to individual or personal aspects of self.” A major source of group identity in many post-colonial nations and powerful authoritarian states from Russia to China is the nation; an identity often actively promoted by the country’s rulers. Information that suggests the nation does not respect human rights is likely to serve as a direct threat to one’s national identity and also to what people think a good patriot should believe.

In the case of international human rights naming and shaming, there is evidence that this backfire effect may be even more noticeable. Research on group-based criticism demonstrates the existence of an ‘intergroup sensitivity effect’; the “tendency for group-directed criticisms to be received in a more defensive way when they stem from outgroup members than when the same comments are made by ingroup members.” This effect appears to be primarily limited to critical information; information that threatens the group identity (people are no more likely to believe messages of praise from ingroups than outgroups), which leads Hornsey and colleagues to argue that group-based criticisms are a “unique subset of persuasive messages in the sense that they directly threaten the (collective) self-concept.”

Of course not all shaming is the same, and many people are also motivated to maintain their self-image as liberal individuals, who also care about human rights, so to some extent there will be a conflict of identities. In simple terms, the backfire effect should be greatest when the threat to the national identity is greatest. I hypothesise that this will occur under the following circumstances:

Source of naming and shaming

Criticism is naturally more likely to appear to be much more threatening to the national identity when it comes from groups whose intentions are perceived to be hostile to the nation than when it comes from those whose intentions are seen as benign or even constructive. In the case of human rights criticism, representatives of foreign groups may appear to be driven purely by interstate competition or long-lasting hatreds. Citizens of the Soviet Union would have been far more likely to perceive criticism of their homeland to be a hostile Cold War tactic pushed by prejudice or an attempt to denigrate the country if it came from the United States than if it came from an ally like Cuba. Indeed Chinese officials often respond to pressure from the United States on their human rights record by

31 ibid. p501
arguing that the West is trying to ‘stop China’s rise’ - criticism is driven by the realist desire for geopolitical superiority rather than any care for the Chinese people’s welfare. This reflects the presumed mechanisms behind the intergroup sensitivity effect; whether the criticiser is seen as having the best interests of the group at heart, or is seeking to attack the group. People generally believe that out-groups will have hostile intentions towards their group, and are therefore more likely to reject their criticism. As such I argue that the perceived hostility of the source of critical information will have the main impact on citizen grievances about human rights conditions.

One main cause of perceived hostility will be the current geopolitical relationship between the source nation or organisation and the target nation; a public will perceive their nation’s geopolitical opponents to be more hostile. When pressure come from geopolitical opponents, members of the public are likely to see the criticism through the lens of international competition - which then heightens the sentiment that the nation is under attack, and defending the national identity is more likely to be a goal. When criticism comes from a neutral or allied source, the threatened national identity figures much less prominently in how publics respond to the new information.

**Hypothesis 1**: Foreign criticism of a country’s human rights will be more likely to reduce public grievances about human rights conditions when it comes from a geopolitical opponent

*Salience of national identity*

We would expect those who identify most strongly with the nation to feel most threatened by critical information. In contrast to political or ethnic identities however, citizens are likely to either identify with the nation, as a patriot, or not at all - and the numbers who do see themselves as proud of their country are generally much more numerous. In the current study in China, 74% of people identified themselves as 8 or above out of 10 on a patriotism scale, with only 10% below the halfway point. And even for those very few who identified themselves as not at all proud to be Chinese, this is primarily a negative identification rather than a positive one. There is no clear ‘opposition supporter’ or ‘democrat’ to identify positively with and defend. As such, when we are talking about the national identity, critical information is much less likely to lead to polarisation than when we are examining political identities.

**Hypothesis 2a**: Criticism of a country’s human rights will be more likely to reduce public grievances about human rights conditions in those with a greater attachment to their nation

This effect should not just work on the individual level however. As overall public nationalism increases - for example when there are international disputes - naming and shaming should be more likely to have a negative effect. At these times the public is not only more likely to be emotionally invested in their nation, but also more likely to encounter flags and other national symbols. As the national identity becomes more salient, people will be more likely to be motivated to defend it - and more likely to create counter-arguments to critical information.

---


33 Hornsey, M.J. (2005)

34 That are the main focus of much of the American political psychology literature (see Cohen, G.L. (2003) for example)
Hypothesis 2b: When public nationalism is high, foreign criticism of a country’s human rights will be more likely to reduce public grievances about human rights conditions.

The national identity should affect beliefs about human rights independent of foreign criticism. Studies have suggested that attitudes to human rights are strongly related to beliefs about the nation. Diaz-Veizdaes and colleagues show that there is a moderate negative correlation between nationalism and belief in the importance of civil liberties. It is not clear however which way this relationship runs - it seems likely that people become more proud of their nation the more they believe it engages in good behaviours like respecting its citizens’ rights. This study gives the opportunity to test whether increases in public nationalism do in fact reduce grievances about human rights.

Hypothesis 2c: When public nationalism is high, public grievances about human rights conditions will be more likely to be lower.

Elites and the nation

If we believe that the backfire effect comes from a defensive reaction to a threat to the national identity, then criticism that does not target the nation as a whole should be less threatening, and therefore have less of a negative effect on grievances. If criticism were to only focus on government elites and their policies in violating human rights and explicitly distinguish between the elites and the nation, then members of the country’s public should be able to reaffirm their belief that their country cares about human rights. They can shift the blame to the elites, who have been decoupled from the nation, and assess the critical message on its own merits.

Hypothesis 3: Foreign criticism of a country’s human rights will be less likely to reduce public grievances about human rights conditions when it explicitly addresses the elites of the country.

Alternative hypotheses

I test these hypotheses against three main alternatives. The first, the ‘informational’ view, is that, as suggested in the literature, foreign criticism about human rights has a generally positive effect on perceptions of human rights conditions. In other words, when publics hear information about condemnation of their country’s human rights situation, their grievances increase. The second, the ‘backlash’ view, shares some similarities with my theory, but argues that all criticism should be rejected equally, no matter the source or the timing. One final view, which does not necessarily conflict with my own theory, is Bush & Jamal’s ‘cues’ approach. They argue that in autocracies citizens interpret support for a regime’s policies as ‘cuing’ support for the regime itself - and therefore will reject or accept the endorsement based on their feelings about the regime. Under this view, no treatments should have any absolute effect, but instead polarise the respondents according to their level of support for the regime. All criticism should make regime supporters see their country’s human rights conditions as better and opponents see the conditions as worse than the control.

Research Method and Case Selection

China

---

I test these hypotheses using an experimental survey of foreign criticism of women’s rights in China. Since the 4th of June 1989, the international human rights regime has singled out China for intense scrutiny. China has faced regular public invocations from Western leaders, parliaments and congresses, media and NGOs, White Papers, United Nations resolutions and even Nobel Peace Prizes over its human rights conditions. However while many have noted the Communist Party tactic of strategically releasing a few prominent prisoners when the eyes of the world’s media have been on them36, evidence of underlying change is less promising. In contrast, in recent years the Chinese Communist Party (CCP) has become notably more intolerant of dissent; tightening laws on civil society, arresting human rights lawyers en-masse, and launching cross-border raids into Hong Kong; called by some ‘China’s new age of fear’37. These efforts have also seemed to have minimal effect on the Chinese public. Despite continued repression and all the international efforts through the 1990s to publicly shame China, in 2001, on the eve of the Falun Gong crackdown, China had the second most positive public perceptions of domestic human rights conditions of the countries surveyed (see Table 1)38. Nationalism has also (perhaps not coincidentally) bloomed in China since 1989. The Patriotic Education campaign has fuelled nationalist outcries against Japan, the United States and France at various points since the 1990s, many erupting into full-scale protests. China is therefore an ideal case to test the impact of nationalism and the effect of shaming.

Table 1: Human Rights Perceptions 2000 - 2004

<table>
<thead>
<tr>
<th>Country</th>
<th>Respect For Human Rights in Country (% say positive - % say negative)39</th>
<th>Total Newsweek/Economist Articles on Human Rights from 1990 - 2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td>66</td>
<td>4</td>
</tr>
<tr>
<td>China</td>
<td>64</td>
<td>53</td>
</tr>
<tr>
<td>Philippines</td>
<td>58</td>
<td>8</td>
</tr>
<tr>
<td>Sweden</td>
<td>50</td>
<td>5</td>
</tr>
<tr>
<td>United States</td>
<td>48</td>
<td>72</td>
</tr>
<tr>
<td>Jordan</td>
<td>46</td>
<td>5</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>44</td>
<td>9</td>
</tr>
<tr>
<td>Egypt</td>
<td>43</td>
<td>18</td>
</tr>
<tr>
<td>Iran</td>
<td>41</td>
<td>20</td>
</tr>
<tr>
<td>Saudi Arabia</td>
<td>41</td>
<td>36</td>
</tr>
<tr>
<td>Uganda</td>
<td>40</td>
<td>7</td>
</tr>
<tr>
<td>Tanzania</td>
<td>38</td>
<td>7</td>
</tr>
<tr>
<td>India</td>
<td>34</td>
<td>29</td>
</tr>
<tr>
<td>Nigeria</td>
<td>25</td>
<td>19</td>
</tr>
<tr>
<td>Spain</td>
<td>24</td>
<td>12</td>
</tr>
<tr>
<td>Japan</td>
<td>23</td>
<td>12</td>
</tr>
</tbody>
</table>

38 This table does not take into account the way in which different country’s citizens interpret these questions on their political circumstances in different ways (see for example King, G., & Wand, J. (2007). Comparing incomparable survey responses: Evaluating and selecting anchoring vignettes. Political Analysis, 15(1), 46-66). However even given these differences the table is surprising, not least in showing that the level of international attention has failed to convince Chinese citizens that their rights are not well respected, however they interpret the question.
39 Those countries in the World Values Survey Fourth Wave, from 2000 - 2004. The question given was: “How much respect is there for individual human rights nowadays in [respondent’s country]? The score is the difference between the percentage that responded A lot of respect or some respect, and those who responded no respect at all or not much respect. World Values Survey Association (2009) World Values Survey 1981-2008 (www.worldvaluessurvey.org).
Despite the powerful censorship apparatus, some of this foreign criticism has reached the public. On one hand there have been ways of reaching around the firewall to access foreign news; from the opening up in the 1980s, information from foreign sources has found its way through in some forms, from those who could understand some English and listen to the Voice of America to more recently those have access to a good VPN to access foreign websites. As an example, the award of the Congressional Gold medal to the Dalai Lama in 2007 was widely celebrated by Tibetans, with a number arrested as a result. They gained knowledge of the award through internet, radio and satellite television sources. On the other side, perhaps concerned about the risk of unfiltered foreign information reaching its citizens, Chinese media have actively reported on instances of foreign criticism of its human rights. Even in the internet-free land of the early 1990s, the state-owned People’s Daily felt obliged to comment on the United Nations draft resolution on human rights in China in March 1993.

Women’s Rights

I focus on the issue of women’s rights for three reasons. International and domestic attention on women’s rights in China has increased dramatically since March 2015, when five feminist activists were arrested and held without trial for a month. Condemnations poured forth from Hillary Clinton and Samantha Power in the US, from the United Kingdom, Canada, the European Union, as well as feminist NGOs from Japan to India. Following extensive public and private attempts to secure the release of the women, they were allowed to leave on bail on the 13th of April. International condemnation, much of which questioned China’s commitments to improving women’s rights, continued unabated throughout the year. A few months later Hillary Clinton called Xi Jinping ‘shameless’ on Twitter, for hosting a forum in New York that celebrated twenty years of...
progress in women’s rights. Chinese state media reported widely on her tweet\(^1\), which was debated on social media. Some women’s rights activists welcomed the comments, saying that they reflecting the frustration many felt with the lack of progress since 1995\(^2\). Other Weibo users criticised Clinton, viewing her comments as deliberate attacks on China rather than arising from any sense of female solidarity\(^3\).

Secondly, if China is a ‘most-likely’ case for a negative effect of criticism, then women’s rights is China is its least-likely version. While Western ideas about democracy and civil rights have been denigrated by the Chinese authorities since the 1990s, women’s rights have been an area in which, historically, foreign intervention has been broadly welcomed - and seen as having generally positive effects. At the end of the Qing Dynasty, foreign missionaries played a major role in ensuring the dissolution of the 1000-year practice of footbinding, working closely with Chinese reformers\(^4\), while since the 1990s funding and ideas from Western feminist groups allowed women’s rights researchers and activists to become far more effective in fighting issues from educational inequalities to domestic violence\(^5\). The CCP has, since its inception, portrayed itself as a liberator of women\(^6\), and as a result party propaganda can ill afford to condemn support for women’s rights.

Finally, research ethics played an important role in the choice of women’s rights. The research environment at the best of times in China prohibits research into sensitive areas of human rights, and this is especially the case in the last few years. There is an obligation to respondents to ensure that they are not asked any questions where their answers may potentially put them at risk. Women’s rights is an issue that, as mentioned, the CCP has been particularly supportive of throughout its history, and is therefore relatively freely discussed in academic institutes and on traditional and social media. This relative freedom also means that social desirability is less of a concern in the survey, and in previous surveys Chinese citizens have generally not been afraid to show dissatisfaction over gender equality - a 2011 Pew survey showed that 46% of Chinese people believed that the country still had improvements to make\(^7\), while in a recent Gender in China survey 73% of women were dissatisfied or extremely dissatisfied with the status of women in the country\(^8\). Social desirability should also not affect how participants respond to the experimental manipulations - the source of criticism or the presence of the Chinese flag.

**Experimental Research Design**

This study uses an online experimental survey, conducted in Chinese during the month of February 2016. The survey was carried out just after the closure of the Beijing Zhongze Women’s Legal Counselling and Service Centre, run by Guo Jianmei, under ‘pressure’ from the Beijing Municipal Public Security Bureau\(^9\). The high profile legal aid NGO had symbolised the growth in women’s rights in China since the 1990s, and the closure again saw international condemnation; Hillary Clinton tweeting: “Women’s rights are human

---

\(^1\) [http://www.globaltimes.cn/content/966853.shtml](http://www.globaltimes.cn/content/966853.shtml)
\(^4\) [http://www.genderwatch.cn:801/detail.jsp?id=305288\&cnID=90050](http://www.genderwatch.cn:801/detail.jsp?id=305288\&cnID=90050)
\(^7\) [http://www.genderwatch.cn:801/detail.jsp?id=305288\&cnID=90050](http://www.genderwatch.cn:801/detail.jsp?id=305288\&cnID=90050)
\(^8\) [http://www.genderwatch.cn:801/detail.jsp?id=305288\&cnID=90050](http://www.genderwatch.cn:801/detail.jsp?id=305288\&cnID=90050)
\(^9\) [http://www.genderwatch.cn:801/detail.jsp?id=305288\&cnID=90050](http://www.genderwatch.cn:801/detail.jsp?id=305288\&cnID=90050)
rights. This center should - I stand with Guo\textsuperscript{56}. The closure of the centre, and Clinton’s response were both reported in the \textit{Global Times}, which suggested that: “it is possible that Clinton was using this women’s rights-related affair to promote her campaign for the upcoming Democratic primary”\textsuperscript{57}. As such, domestic awareness of women’s rights issues and the international interest was comparatively high.

The survey was conducted with an online sample of 1200 Chinese netizens, using the Qualtrics survey provider and Qualtrics’ panel providers in China. For age and gender, respondent numbers were weighted to match the distribution in the overall population\textsuperscript{58}. With the exception of these demographics, the sample more closely resembles the online population - richer, more well-educated and urban- than the overall Chinese population\textsuperscript{59}, but was drawn from almost all provinces and walks of life. While we should not neglect these demographic differences, the better-educated urban population are likely to be more politically active than those in China’s poorer rural areas. The online population (almost 50\% of the population in mid-2015)\textsuperscript{60} are arguably the most likely to pick up on foreign comments about China, and some have argued that amongst the middle class in China, political and civil society participation is now more likely to be online\textsuperscript{61}. Beijing’s censorship apparatus is specifically designed to prevent mobilisation within this social group\textsuperscript{62}. Finally, due to the sensitive nature of the topic, a face-to-face nationally representative survey would also have faced ethical problems and also less likely to reveal honest answers than an anonymous online questionnaire.

This methodological approach has been used rarely in the international human rights literature, and even studies of public opinion have generally relied on large cross-national cross-time datasets. The experimental design allows me to isolate instances of foreign naming and shaming of China’s human rights conditions and examine directly how small variations in these criticisms affect the public. The experiment involved randomly exposing respondents to foreign criticism of women’s rights in China, randomly assigning both the source of criticism and manipulating the level of nationalism, in a 3 x 2 factorial design\textsuperscript{63}.

\textbf{Source of criticism}

Respondents were assigned to one of three groups, a control group, who were given no prompt, and two treatment groups who were asked to read a short paragraph taken from a recent news item, as follows:

\textbf{Treatment 1}: Yesterday a United States spokeswoman criticised China’s women’s rights conditions. She said: “The Chinese government must improve the rights of women in China”

\textsuperscript{57} http://www.globaltimes.cn/content/966853.shtml and http://opinion.huanqiu.com/shanrenping/2016-02/8487620.html
\textsuperscript{58} These were the only weightings that could be reasonably made through the survey panel.
\textsuperscript{59} Discussed in Appendix 1
\textsuperscript{60} See China Internet Watch (2015) \textit{China Internet Statistics}. Accessed at: https://app.box.com/s/pk6wyre4unf0i4n071stnqet4qghlnr
\textsuperscript{61} Yang, G. (2009) \textit{The Power of the Internet in China: Citizen Activism Online}. Columbia
\textsuperscript{62} King, G., Pan, J., & Roberts, M. E. (2013). How censorship in China allows government criticism but silences collective expression. \textit{American Political Science Review}, 107(02), 326-343
\textsuperscript{63} The randomisation procedure was successful, as discussed in appendix 2, Table 8. There are no statistically significant differences in demographic variables or in the pre-treatment attitudinal questions, with the exception of education, which is minimally significant at the 10\% level.
Treatment 2: Yesterday an African Union spokeswoman criticised China’s women’s rights conditions. She said: “The Chinese government must improve the rights of women in China”

The United States is currently the main geopolitical opponent of China⁶⁴, and has been historically in varying degrees since the Communist Revolution in 1949. The African Union is at worst a neutral actor for the Chinese people, and at best a long-lasting geopolitical ally. Again, since the Maoist period, Beijing has portrayed itself as the leader of the developing world, providing assistance to anti-imperial independence causes. China received broad support from African countries in its fight against United Nations resolutions on its human rights in the 1990s, and even in recent months the People’s Daily described the relationship as “friendly” and a “a community of mutual support”⁶⁵.

Nationalism

The second condition involves the manipulation of the levels of nationalism of the respondents. The control received no prompt, while the treatment group received a small Chinese flag (measuring 1” x 0.5”) placed in the top left-hand corner of the screen when answering outcome questions. Scholars in political psychology have used national flags as a way of increasing the salience of the ‘nation’ in respondents. Schatz and Lavine argue that national symbols “uniquely accentuate citizens’ identification as national members”⁶⁶, and scholars have shown that even subliminal exposure to national flags can manipulate political attitudes and decisions⁶⁷. The Chinese flag is also frequently associated with times when China is engaged in international disputes - it was, for example almost ubiquitous in the mass-anti-Japan protests held in 2012.

Table 2: 3 x 2 experimental design

<table>
<thead>
<tr>
<th></th>
<th>United States Criticism</th>
<th>African Union Criticism</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Flag</td>
<td>Control</td>
<td>No nationalism</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hostile source</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-hostile source</td>
</tr>
<tr>
<td>Flag</td>
<td>Nationalism only</td>
<td>Nationalism</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hostile source</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-hostile source</td>
</tr>
</tbody>
</table>

After receiving the treatments, respondents then given some distractor questions⁶⁸, and were then questioned on their beliefs about women’s rights in China. For the main outcome variable respondents were asked for their level of agreement with a statement that women’s rights are not good enough in China (on a four-point scale). Three questions, for robustness, then test different measures of respect for women’s rights in China. The second block measured respondents’ attitudes towards women’s rights, and were combined into one ‘attitude’ variable⁶⁹. These questions are listed in full in Table 3,

---

⁶⁴ Alongside Japan - although human rights criticism rarely comes from Tokyo
⁶⁵ http://en.people.cn/n/2015/1207/c900000-8986939.html
⁶⁸ Two questions about leisure activities, not related to the topic of the study.
⁶⁹ These responses were highly correlated, so show internal consistency (Cronbach’s alpha = 0.7815)
and the distribution for the main outcome variable in the control group is shown in figure 1 below:

Figure 1: Distribution of main outcome variable in control group

Finally, in a survey like this it is difficult to see how grievances about women’s rights might translate into attempts to do something about them. I therefore included a question asking whether respondents would be willing to sign a petition calling for improvements in women's rights (of course willingness to sign does not mean people actually would sign). Petitions are a historically-common and state-recognised way in China in which social grievances have been presented to local and central governments - there were 6.4 million petitions between January and October 2013\(^1\). I normalise all of these outcome variables, to measure change from the control group.

I also make an initial test of my hypothesis 3. For those four groups who received criticism earlier in the survey, at the end I include a further sentence that continues the news story. It reads: “The spokeswoman continued: “Although the Chinese people have continuously struggled for women’s rights, it is the Chinese government that has not ensured women’s rights are good enough in recent years”\(^2\). Respondents were then asked again about their grievances on women’s rights in China.

How do the hypotheses map onto these experimental tests? From H1 I expect that this effect will be mainly limited to those who received criticism from the United States - and that those who received criticism from the African Union will show little negative effect - and markedly less than the United States condition. H2a suggests that this effect will be largest in those with higher pride, while H2b suggests that it will be larger in the conditions when respondents are also given a Chinese flag. According to H2c this ‘nation’ effect will remain even without criticism. Finally, from my H3, I would expect that the groups who received criticism, once they receive the second passage that emphasises the criticism is only directed at the elites, will have stronger grievances about women’s rights than before this second passage.

\(^{70}\) Other summary statistics for the control groups are included in the appendix.

\(^{71}\) http://www.chinadaily.com.cn/china/2014-06/16/content_17589189.htm
Table 3: Response questions

<table>
<thead>
<tr>
<th>Questions</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Grievances</strong></td>
<td></td>
</tr>
<tr>
<td>“At present China’s women’s rights’ situation is not good enough”</td>
<td>1(strongly agree) - 4 (strongly disagree)</td>
</tr>
<tr>
<td><strong>Supplementary Perceptions</strong></td>
<td></td>
</tr>
<tr>
<td>“Are women’s rights respected in China?”</td>
<td>1(not at all) - 4(high respect)</td>
</tr>
<tr>
<td>“In most aspects, China has already done enough to improve women’s rights”</td>
<td>1(strongly agree) - 4 (strongly disagree)</td>
</tr>
<tr>
<td>“In China, men and women’s education opportunities are not equal”</td>
<td>1(strongly agree) - 4 (strongly disagree)</td>
</tr>
<tr>
<td><strong>Attitudes</strong></td>
<td></td>
</tr>
<tr>
<td>“University education is more important for boys than girls”</td>
<td>1(strongly agree) - 4 (strongly disagree)</td>
</tr>
<tr>
<td>“Men belong in public life, women belong in the home”</td>
<td>1(strongly agree) - 4 (strongly disagree)</td>
</tr>
<tr>
<td>“When jobs are scarce, men have more right to work than women”</td>
<td>1(strongly agree) - 4 (strongly disagree)</td>
</tr>
<tr>
<td><strong>Behaviours</strong></td>
<td></td>
</tr>
<tr>
<td>“Are you willing to sign a petition calling on the government to put in place policies that improve women’s rights?”</td>
<td>1(strongly agree) - 4 (strongly disagree)</td>
</tr>
<tr>
<td>“Are you willing to sign a petition calling on the government to not put in place policies that improve women’s rights?”</td>
<td>1(strongly agree) - 4 (strongly disagree)</td>
</tr>
<tr>
<td><strong>Follow-up question</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Grievance:</strong></td>
<td></td>
</tr>
<tr>
<td>“At present China’s women’s rights’ situation is good enough”</td>
<td>1(strongly agree) - 4 (strongly disagree)</td>
</tr>
</tbody>
</table>

**Results**

I find little support for either the informational or the backlash hypotheses, that overall criticism has an aggregate positive or negative effect on grievances about women’s rights (when excluding nationalism, average treatment effect (ATE) = -0.12, p=0.19). However when we split up the criticism by source there is a notable impact. When criticism comes from the United States (in the absence of any nationalism treatment), the ATE is -0.244 (p=0.02); meaning that people hearing criticism from this source have significantly lower grievances about women’s rights. Grievances following criticism from the African Union on the other hand are statistically indistinguishable from no criticism (ATE= -0.003, p=0.98). This provides strong support for my hypothesis 1; that criticism from a geopolitical opponent, the United States, has a much more negative impact on grievances in comparison to a neutral foreign actor, the African Union (ATE= -0.247, p=0.021). Figure 2 demonstrates these effects.
Figure 2: Influence of treatments on grievances over women’s rights (higher score means higher grievances, dotted line is the control)

Nationalism also appears to have an impact on grievances. I find strong support for my hypothesis 2c, as in the absence of criticism, those respondents who saw a Chinese flag had significantly lower grievances (ATE = -0.239, p=0.021). The flag treatment effect holds across all conditions (ATE = -0.144, p=0.018). However in the criticism conditions, while still having a positive effect, the results were non-significant (for the US ATE= -0.109, p=0.296; for the AU ATE= -0.082, p=0.444). This suggests that the interactive effect I describe in hypothesis 2b, where nationalism makes criticism even more counterproductive does not hold. Instead, nationalism has an additive effect to the negative impact of criticism from the United States; and the ATE for this additive effect (US criticism plus nationalism) is -0.354 (p<0.001). The backfire effect, as predicted in hypothesis 2a, is strongest in those with higher pride: but only for US criticism in the non-nationalism condition (for the interaction with pride F=4.69, p=0.031) but not for criticism from the AU or when the Chinese flag was primed.
These results also, to some extent, translate into behaviours - the willingness to sign a petition for women’s rights. US criticism makes respondents less willing to sign the petition in comparison to the control (ATE = -0.268, p=0.015) and in comparison to AU criticism (ATE=-0.337, p=0.002), while AU criticism itself has a non-significant positive effect (ATE=0.069, p=0.50). Exposure to the Chinese flag has a minimal and non-significant effect in all cases.

I find surprisingly strong support for my hypothesis 2c; that when criticism only targets the elites it will be less likely to reduce grievances about women’s rights. A secondary prompt saying that the initial criticism was targeted at the Chinese government not the people significantly increased grievances with women’s rights in comparison to full criticism (but only in the US condition), as shown in Figure 4. With US criticism alone, the extra prompt increased grievances with ATE=0.233 (p=0.004) against the full criticism, while when the US criticism was accompanied by a flag, the prompt increased grievances with ATE=0.461 (p<0.0001). Indeed, in this condition the prompt significantly increased grievances about women’s rights against the control group (ATE=0.172, p=0.029).

\[72\] This analysis is somewhat complicated by the fact that the wording of the question following the second prompt was slightly different from the first. In the control group, the mean reversed score for the first question: “At present China’s women’s rights’ situation is not good enough” is 2.57, while the mean score for the second question: “At present China’s women’s rights’ situation is good enough” is 2.65, a difference of 0.075. I standardised the second question for all groups by adding the difference to each answer.
Robustness checks

It is possible that the fall in grievances displayed above may come not from changes in respondents’ beliefs about how well China respects women’s rights, as suggested in my theory - but instead from reducing how much respondents care about women’s rights. The below Figure 5 of composite attitudes towards women’s rights suggests that this is not the case. If anything, criticism from the United States strengthens support for women’s rights in China (for non-nationalism, ATE=0.155 (p=0.11); for nationalism, ATE=0.129 (p=0.19)). Criticism from the African Union significantly strengthens support ((for non-nationalism, ATE=0.194 (p=0.048); for nationalism, ATE=0.231 (p=0.019)).
I also test three other measures of ‘respect’ for women’s rights in China. Responses on whether China respects women’s rights were very similar to the main dependent variable, albeit with slightly lower effect sizes, as shown in appendix 3. The other two measures however showed slightly different and weaker effects. The question on whether the Chinese government has done enough for women’s rights showed that the US criticism and nationalism treatments both had significant, but lower negative effects (ATEs around 0.2), but their combination was non-significant (ATE = -0.09). The final question was issue-specific, and asked whether girls’ education opportunities were lower than boys’ in China. While the pattern was the same as the main dependent variable, the effect sizes were much smaller, with only the combination of US criticism and nationalism leading to significant increases in perceptions of equal education opportunities (ATE = -0.207, p=0.037). The mixed results could be explained by the fact that these issues (past government performance and education opportunities) were not directly addressed by the criticism, and responses to them bring in external factors about the education system and previous government behaviour rather than current grievances about women’s rights.

The findings about reduced willingness to sign petitions could potentially be explained not just by reduced grievances about women’s rights- but by a lower willingness to sign petitions in general (for example if foreign criticism reduced trust in government ability to respond to the public). To control for this I created a measure for the difference between willingness to sign a petition for women’s rights, and the willingness to sign a petition against women’s rights. The results are similar, with smaller effect sizes. The ATE for US criticism without nationalism is -0.145 (p=0.166) against the control, while for US criticism with nationalism was -0.221 (p=0.03)73.

73 See appendix 3 for full results
Looking at interactive effects, I find that in the US criticism condition where there is no nationalism, as expected, the results are mainly driven by citizens who are satisfied with the regime\(^{74}\) and distrust the US government\(^ {75}\) - but also those who are university-educated\(^ {76}\) and have a positive attitude towards the United States as a whole\(^ {77}\). These effects do not exist for criticism from the African Union and generally disappear in the nationalism condition. I also find no interactive effects by awareness of international news - this suggests that those who have been more aware of the recent criticism of China by Hillary Clinton were not biased in either way by this knowledge There were no other significant interactions with any other pre-treatment questions, including by age or gender.

**Discussion**

My findings show that criticism from the United States has a backfire effect; significantly reducing grievances about women’s rights in China. Criticism that does not come from a geopolitical opponent, the African Union, has no negative effect on grievances. This suggests that publics evaluate the perceived hostility of the source of naming and shaming information when forming grievances about human rights. The United States, a geopolitical opponent, could be perceived to have ulterior motives in criticising women’s rights in China, motives that are notably absent in a neutral or allied actor like the African Union. Reacting defensively against this perceived hostility, members of the public strengthen their beliefs that their country respects women’s rights, possibly by thinking through counterarguments.

Findings about nationalism lend cautious further support to this argument. In the absence of foreign criticism, exposure to a Chinese flag decreased grievances about women’s rights. This demonstrates that increases in symbolic nationalism, likely to appear around times of international disputes or national holidays, have a causal effect on lowering grievances about human rights, rather than merely a correlation. The results also suggest that around these times, criticism will have more of a negative effect on grievances. However this effect appears to be additive rather than interactive, in contrast to my predictions. This finding may come from the fact that foreign criticism is itself likely to raise nationalism in members of the Chinese public. As more people become highly nationalistic as a result of the criticism, the treated symbolic nationalism will itself have diminishing returns, which may balance out any positive interactive effects.

*Who is more likely to be affected?*

As well as those with higher patriotism, my results also generally support those of Bush & Jamal; that respondents who favour the regime will respond more negatively to criticism of their government’s policies\(^ {78}\). My results fit this conclusion; that those who are satisfied with the regime overall are less likely to agree with outside criticism of the regime’s policies. However Bush & Jamal’s theory suggests that regime opponents should accept criticism just as supporters reject it - yet my results show that those who are not fully satisfied with the regime instead have little positive response, fitting more the conclusion that foreign pressure leads to an overall negative reaction rather than polarisation. There seems to be a stronger negative psychological reaction from citizens who are satisfied with their regime, and proud of their country, that goes beyond simply rejecting the criticism - and that does not appear in the opposite fashion from those dissatisfied with the regime.

\(^{74}\) F=4.15, p=0.04  
\(^{75}\) F=7.89, p=0.005  
\(^{76}\) F=5.19, p=0.023  
\(^{77}\) F=6.33, p=0.012  
The negative impact on regime supporters also only comes in the case of a hostile foreign source, but not when the source is a neutral actor. Chinese regime supporters may still disagree with criticism from the African Union - but it does not push them to become more satisfied with women’s rights - this effect only appears when the source is the United States.

Perhaps surprisingly, it is not just conservative CCP supporters who react negatively. They are highly patriotic, but are also generally well educated and internationally oriented. These are groups more commonly associated with a more liberal demographic. And indeed in the control, these groups are those who are significantly more likely to believe that women’s rights are not good enough in China at the moment. In other words, those who are most affected by the US criticism are those who are satisfied with their regime, and proud of their country - but are more liberal and more likely to believe that women’s rights need to be improved. I split the outcome variable into a binary one for those who generally believe women’s rights to be good enough versus those who generally believe women’s rights are not good enough. US criticism makes respondents significantly more likely to believe criticism is ‘good enough’ than ‘not good enough’ (against the control, ATE=0.144, p=0.004). Hostile criticism not just strengthens the satisfaction of those already pleased with their regime, but also actively shifts people across the divide, from dissatisfaction about women’s rights to satisfaction.

Seen in this light, the results are particularly worrying - as US criticism does not just strengthen the resolve of people who would not necessarily believe women’s rights are a problem in any case- but persuades people who might otherwise have opposed government policies to instead support them. These people are generally also proud of their nation and regime supporters; who are proud of their nation and appear overall to be satisfied with the regime’s performance on other issues more important to them than women’s rights - but believe that it can do better with this specific issue - that is, until the issue is criticised by a hostile geopolitical opponent.

One final finding that lends support to my theory of perceived hostility is the interaction between pre-treatment trust in the United States government and criticism from the United States. I argued that negative responses to criticism are driven to some extent by the belief that the United States is a hostile power with ulterior motives. I do not ask this question directly, but find that those who distrust the US government (5 or below on a 1-10 scale) are significantly more likely to reduce their grievances about women’s rights having heard criticism from the US (by a value as high as ATE=0.55: see figure 6).

79 These results are surprising on the face of it. On one hand more liberal respondents should hold a more ‘liberal’ identity and stronger existing beliefs that women’s rights need to be improved- which they should be interested in preserving even in the face of a conflicting national identity. One the other hand, people should have to value women’s rights in some form first to feel threatened by criticism of those rights in their country
Figure 6: Interaction of trust in US government and United States criticism on grievances about women’s rights, when no nationalism

It is worth noting that almost all of these interactive effects disappear once people are exposed to a Chinese flag. In this condition, people more or less uniformly reject US criticism and lower their grievances about women’s rights. This suggests that one main impact of increasing symbolic nationalism is to bring everybody together around one position of rejecting hostile foreign criticism. In times of conflict or around anniversaries, criticism is likely to have a particularly strong effect on those who are less satisfied with the regime or who are less patriotic. This supports the psychological literature on symbolic nationalism, which argues that exposure to national symbols converges people’s attitudes towards that of a ‘typical citizen’80 - in this case one who rejects foreign criticism of the nation.

Perhaps the most striking results are on the impact of criticism that explicitly targets the Chinese government elites over women’s rights, as opposed to the whole country. This criticism aims to break down the link between the governing CCP and the nation itself, to reduce the perceived threat to the nation from foreign shaming. I find that when United States criticism of women’s right in China is supplemented to say that it targets only the CCP elites, and not the people, the counterproductive effects vanish. And in conditions where nationalism is high, criticism from the United States elicits a positive response from the Chinese public, increasing grievances about women’s rights.

My results imply that when the United States criticises China as a whole nation, the criticism and the presence of the Chinese flag together evoke a defensive nationalism, which leads members of the public to reject negative comments about women’s rights. However when the criticism explicitly splits up the regime and the public, my findings

---

suggest that nationalism becomes directed inwards at self-improvement. More widely, this proposes an interesting extension to the psychological literature on group-based criticism and attitude change in a political context. It implies that not only does the nature of the source matter, but that when we talk about international human rights pressure, we also need to take into account the relationship between the public and their leaders; the two potential targets of criticism.

The effects are substantial. When United States criticism is accompanied by nationalism, people’s beliefs that China’s women’s rights are good enough increases by 0.354 (against the control). This is not far short of the difference between men and women in the sample (ATE=0.391). When the scale is translated into those who believe rights are good enough versus those who believe they are not good enough (regardless of strength of belief), US criticism makes people 28.6% more likely to say that they believe rights are good enough. If the national identity is primed, US criticism makes people 37.7% more likely, while for those with higher than average levels of patriotism, US criticism on its own makes people 42.3% more likely. Lower grievances also make people less willing to sign petitions for women’s rights. Criticism from the United States decreases the likelihood of signing a petition by 0.268, a difference that is over half that of the gender gap (ATE=0.47). It makes people 32% less likely to say they will sign rather than not sign a petition.

Criticism from the African Union on the other hand has no impact on either grievances or willingness to sign. AU criticism in fact has positive impacts on attitudes towards women’s rights - in that it makes people more likely to believe in rights for women. Again this impact is sizeable - criticism of women’s rights in China by the African Union liberalises attitudes by 0.231 - almost identical to the difference in beliefs between men and women. Support for women’s rights overall (on an adjusted scale) is around 36.1% higher in this condition. This is a wholly unexpected finding, given the well-established difficulties in changing attitudes in the psychological literature, especially from outgroups, and is something to be analysed further in future.

**Conclusion and Implications**

My findings demonstrate the importance of the source and of the timing of international pressure on human rights. Criticism that comes from a geopolitical opponent, and that comes at a time in which the target country is involved in an international dispute will be much more likely to have a counterproductive ‘backfire’ effect on public grievances about human rights. In the China case this suggests that intense pressure from the United States on China’s human rights since the 1990s may have even contributed to the public perception that the CCP was respecting human rights, and may even have contributed to the failure to achieve long-lasting change.

The prognosis is not completely negative however. Criticism from a neutral source not only has no negative effects on grievances but also may increase how much members of the public value human rights. Criticism that does come from a geopolitical opponent in times of heightened nationalism does have the potential for positive changes; increasing grievances about human rights when it explicitly targets elites and not the nation as a whole. The question of course is whether the kind of message used in the survey is likely to reach the public when censorship is high - in authoritarian countries, while state media

---

\(^{81}\) Percent not percentage points. The difference is 14 percentage points

\(^{82}\) A difference of around 10 percentage points
may often report on US criticism of its human rights, it will generally seek to frame the message as as favourable a way as possible. However the theory suggests that it is not just the phrasing of the criticism that matters, but the relationship between elites and public that it targets. In other words, if the elites are not seen to be properly representative of the nation as a whole or not protecting the country adequately, then foreign criticism may have a positive effect. Future work could test this further by examining the impacts of foreign criticism at times when the government appears to be acting aggressively to defend the nation - and times where it appears to be giving in or making concessions to foreign powers.

The results also face concerns about external validity; whether the results only apply to the peculiar issue of women’s rights in China. Firstly, China is a place where in recent years criticism from the United States about human rights in China has been linked heavily in state-owned media with a purported desire to prevent China’s rise. As noted, this makes it something of a most-likely case for this kind of analysis. However this kind of propaganda is used regularly in many authoritarian regimes, from Russia to the Middle East. As such a lot of those countries that are most often targeted for human rights pressure (and where it most often fails) are ‘most likely cases’. For those cases where this kind of rhetoric is used less often, the impacts are less clear. It seems likely that in these countries the belief that criticism from a geopolitical opponent is driven by interstate competition, whereas criticism from allies may come from more benign intentions would not be purely a result of government propaganda. While the effect size might be smaller, the mechanisms involved should be the same. On the issue of women’s rights, this concern is less damaging. There is no clear reason why respondents should respond more negatively to criticism on the issue of women’s rights than any other rights issue. And indeed, as noted above, given the level of support afforded by the Chinese government on women’s rights, we might in fact expect Chinese citizens to be relatively open to criticism, in contrast to issues more heavily linked to Western attacks on China in state media.

One other concern might be on the nature of the experimental survey itself and its stylised form. Can we really argue that this kind of manipulation truly represents exposure to foreign criticism over time, and that an immediate snap judgement about the state of women’s rights in China represents a long-term grievance about human rights conditions? Probably not - but this kind of focussed experimental study helps to direct attention to the mechanisms involved in public reactions to naming and shaming, and provides much-needed building blocks to larger scale qualitative studies. While the timescales involved in forming a response are short, the effects are large, especially since the prompt itself is so small. In China, the public will have built up a long history of western criticisms of their human rights behaviour, which presumably would multiply the effects from one prompt. Future work should look at long-term impacts of repeated exposure to these kinds of comments, and whether they have led to any behavioural changes.

---

83 For example see Jetschke, A. (1999)
84 https://www.washingtonpost.com/world/europe/russias-anti-us-sentiment-now-is-even-worse-than-it-was-in-soviet-union/2015/03/08/b7d534c4-c557-11e4-a188-8e4971d37a8d_story.html
87 Admittedly this is difficult in the current climate in China
A final and linked issue is that after hearing criticism respondents may just be giving an immediate angry reaction, defensively refuting criticism - but do not truly believe that women's rights are better as the theory suggests. This would mean that the study is not measuring changing views, but brief annoyance. While distractor questions were used, this objection is again difficult to refute definitively without long-term analyses or more detailed interview work. We would also need to theorise how an angry reaction as opposed to a lower grievance would translate into different responses in the public sphere and civil society. The more complex manipulations in the study do suggest that this is not the whole story. If people were just giving superficial 'reactions' to criticism then we might expect respondents to also give some form of short-term angry response to criticism from the African Union - yet this is entirely missing, even in those with highest national pride. Neither did criticism evoke a defensive reaction on attitudes towards women's rights - instead of superficially dismissing the issue, it made people appear to care more about women's rights.

The study redirects attention to how a target country’s public respond to the international efforts to change behaviour on human rights, and challenges commonly-held views about the relationship between international pressure ‘from above’ and domestic pressure ‘from below’. According to these views foreign shaming, threats, and sanctions impose costs directly on governments, but also indirectly by working in tandem with and providing support to domestic movements. Some scholars have advocated for a ‘comprehensive approach’, which calls for a combination of these direct and indirect efforts - attacking the elites from above as well as encouraging the inculcation of broader public norms. The results here however challenge the assumption that international pressure will invariably have a positive impact for domestic groups, and that foreign and international efforts will work in tandem to change government behaviour. Instead they suggest that under some conditions, foreign pressure will make domestic groups’ efforts more difficult, by reducing the likelihood that members of their public will support their cause and making longer-term normative changes less likely.

Indeed to make the comprehensive approach even more troubling, it seems that what works from above may even inhibit work from below. The typical view is that top-down pressure relies on leverage - and that state compliance will be more likely when the source is powerful and the target weak. As Keck and Sikkink say for China, "what is often missed in the debate over the apparent 'failure' of human rights policy in China is that virtually none of the classic military and economic levers exist." Leverage is essential to promoting change - and indeed, even on the issue of women's rights in China, many NGOs and feminist activist groups have called out to the United States to publicly criticise the Chinese government. In many cases this kind of top-down pressure is clearly effective for encouraging short-term concessions.

This study points to a potential contradiction between leverage and public persuasion however. In some cases, when the pressure comes from a threatening source, this more effective kind of ‘top-down’ pressure may actually make bottom-up societal pressures less likely. Sometimes the public impacts of international pressure will coincide with the top-

---

91 For example http://www.rfa.org/english/news/china/feminists-07062015134642.html
down behavioural impacts; for example in US allies. At other times however, greater top-
down leverage will be the most counterproductive. In the case of China or Russia, public
shaming of human rights will be most successful in the public when it comes from non-
Western sources, those who do not appear to have any ulterior motives but arguably also
the least leverage in international relations.

This reiterates that to properly understand whether international human rights pressure is
successful or not we need to recognise the geopolitical link between the source and the
target of the pressure. The literature on state responses to international pressure has
tended to treat the international system as a homogenous entity. Reviews of the literature
on economic sanctions,\textsuperscript{92} naming and shaming,\textsuperscript{93} and the impact of foreign pressure on
the collapse of authoritarian regimes\textsuperscript{94} rarely mention how the source or timing of the
pressure might matter, and those that do focus mainly on the concept of leverage\textsuperscript{95}. One
exception is Levitsky and Way, who argue that leverage has certainly affected how
vulnerable states have been to Western democratising pressure. However for them
leverage is not enough. Essential to this vulnerability is the \textit{linkage} of the target country to
the West - in other words when there are dense ties (for example geopolitical and
economic) between the target and the West, the costs of resisting the democratising
pressure rise.\textsuperscript{96} By focussing on a different kind of linkage, this study is designed, in part,
to follow their example.

\textsuperscript{92} For example: Drezner, D. W. (2011). Sanctions sometimes smart: targeted sanctions in theory and
\textsuperscript{94} Geddes, B. (2002, August) The Effect of Foreign Pressure on the Collapse of Authoritarian Regimes.
In \textit{Annual Meeting of the American Political Science Association, Boston}.
\textsuperscript{95} See Cardenas, S. (2004)
APPENDICES

Appendix 1

Survey description
The survey used the Qualtrics survey provider and Qualtrics’ panel providers in China. Qualtrics providers hold online panels of netizens throughout China, who participate in surveys for points - and the chance to gain rewards if enough points are gained. For each new survey, Qualtrics sends an email to a random selection of those in the panel, saying “You are invited to participate in a general opinion survey!”, ensuring they stay within a certain demographic quota. The email recipients can then choose whether to participate in the survey or not. The survey was carried out between the 5th of February and the 26th of February 2016, with the majority of the data collected between the 13th and the 17th of February. This was due to the need for two ‘soft launches’ to test the survey at the start and the longer time it took to elicit responses from harder-to-reach demographics towards the end. Princeton University’s Institutional Review Board approved the survey, with protocol #0000007369.

Data quality
The survey and panel provider Qualtrics uses various techniques to improve respondent quality. These include minimising the number of surveys each panel member can take per month, to reduce fatigue and overuse; and regular verification of the panel pool demographics. The survey itself also used tools to ensure the quality of the data. ‘Attention check’ questions filtered respondents out who may have been randomly clicking boxes; while those who completed the survey in an unreasonably short time (one third of the median response time) were also excluded. In total, out of 1723, 523 people were removed from the results by these measures. 11 did not give consent, 228 failed the attention check, and 12 completed the survey below the minimum cut-off time. 269 failed to complete the survey at all. This left a total of 1200 different respondents who completed the survey satisfactorily.

Data description
The tables and figures below describe the summary statistics for the respondent demographics. As shown, the gender and age proportions broadly coincide with the broader Chinese population, as taken from the official 2013 census (reported in 2014). While I did not measure income levels, the income levels in the panels from which the respondents were selected are generally higher. Figure 2 also demonstrates that the survey draws its members primarily from the urban population. This suggests that the overall demographics from this sample are far more urban, educated and wealthy than the population at large - and more likely to be CCP members - perhaps reflecting the higher education levels. I also use the 2015 China Internet Watch report to compare to the online population, which is also more urban and highly educated. However since this also includes a large proportion of under-18s, excluded from my study, comparisons are difficult.

---

97 http://www.stats.gov.cn/tjsj/ndsj/2014/indexeh.htm. Note that this census includes respondents from 6 and above rather than the 18 and above in my survey
Table 1: Demographic Proportions

<table>
<thead>
<tr>
<th></th>
<th>Survey</th>
<th>2013 Census</th>
<th>2015 CNNIC survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women</td>
<td>49%</td>
<td>48.8%</td>
<td>44.6%</td>
</tr>
<tr>
<td>Han</td>
<td>97.1%</td>
<td>91.51%</td>
<td>-</td>
</tr>
<tr>
<td>Current CCP members</td>
<td>26.6%</td>
<td>6.46%*</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 2: Age

<table>
<thead>
<tr>
<th>Age</th>
<th>Survey</th>
<th>Census</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-24</td>
<td>0.163</td>
<td>0.112</td>
</tr>
<tr>
<td>25-34</td>
<td>0.233</td>
<td>0.203</td>
</tr>
<tr>
<td>35-44</td>
<td>0.236</td>
<td>0.217</td>
</tr>
<tr>
<td>45-54</td>
<td>0.193</td>
<td>0.194</td>
</tr>
<tr>
<td>55+</td>
<td>0.175</td>
<td>0.274</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th>Internet</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;19</td>
<td>0.26</td>
</tr>
<tr>
<td>20-29</td>
<td>0.312</td>
</tr>
<tr>
<td>30-39</td>
<td>0.232</td>
</tr>
<tr>
<td>40+</td>
<td>0.191</td>
</tr>
</tbody>
</table>

Figure 1: Age by proportion

http://www.china.org.cn/china/2015-06/29/content_35939304.htm