ARTICLE

In Search of Self-Censorship

Xiaoxiao Shen* and Rory Truex*

Princeton University, Princeton, NJ, USA
*Corresponding author. E-mail: xiaoxiao@princeton.edu and rtruex@princeton.edu

Abstract
Item nonresponse rates across regime assessment questions and nonsensitive items are used to create a self-censorship index, which can be compared across countries, over time and across population subgroups. For many authoritarian systems, citizens do not display higher rates of item nonresponse on regime assessment questions than their counterparts in democracies. This result suggests such questions may not be that sensitive in many places, which in turn raises doubts that authoritarian citizens are widely feigning positive attitudes towards regimes they secretly despise. Higher levels of self-censorship are found under regimes without electoral competition for the executive.

Citizens living under authoritarian rule are thought to engage in preference falsification, voicing support for their governments in public but harboring disdain in private. In aggregate, this behavior makes it difficult to gauge the level of support for a given regime (Kuran 1991; Kuran 1997). When we observe positive attitudes towards authoritarian governments, it may be because that support is authentic, or it may be because citizens are adept at 'living within the lie', to use Havel's (2009) phrase.

Despite the influence of the concept, the empirical evidence for preference falsification in authoritarian systems is limited. In China, Jiang and Yang (2016) show that citizens' assessments of the regime using sensitive and nonsensitive questions diverge after a local political purge, which is taken as evidence of falsification.[1] In rich ethnographic accounts, Wedeen (2015) provides detail of Syrian citizens behaving 'as if'– engaging in regime-affirming rituals and displays, despite being more critical in private. Other studies reach the opposite finding. Frye et al. (2017) conduct a list experiment to assess the popularity of Vladimir Putin and find little evidence of falsification, concluding that, 'Putin's approval ratings largely reflect the attitudes of Russian citizens'. Combined, these single-country studies suggest that preference falsification may exist, but that it is likely context dependent.

It is difficult to directly measure falsification, but we may readily measure a related behavior – self-censorship. This article proposes a simple test for assessing the likelihood of self-censorship at scale, using existing cross-country survey datasets. The test involves calculating the difference in item nonresponse rates ('Don't Know' and 'No Answer' responses) for regime assessment questions and nonsensitive political questions (Berinsky 1999; Berinsky 2004; Berinsky and Tucker 2006). This gives a rough indicator of the number of citizens that choose reticence

---

[1]Recent experimental studies in China show that citizens do not give more politically correct answers when told a survey was affiliated with the Chinese Communist Party (CCP), which suggests that the influence of 'political wariness' on survey responses may be limited in that context (Lei and Lu 2017).

when they assess the regime (Kuran 1997). This index can be replicated across samples and population subgroups, allowing us to directly compare different authoritarian systems.

We implement this test using data from the most recent waves of the World Values Survey (WVS). While the WVS is not beyond criticism (King et al. 2004), it represents the largest-scale survey effort in human history, covering 100 countries over three decades. The WVS core questionnaire contains a set of direct regime assessment measures, including questions on confidence in government, democracy and respect for human rights. The WVS has been implemented in fifty-nine country-year samples in political systems that would be considered authoritarian by conventional standards.

Our core findings are as follows. First, a simple bivariate regression shows that the self-censorship index is indeed associated with regime type. In general, citizens living under authoritarian rule refuse to answer questions about confidence in government, democracy and human rights at slightly higher rates than their counterparts in democracies.

Secondly, there exists important variation in self-censorship within the authoritarian world. The standard deviation of the index in authoritarian country-years is substantially greater (0.031 vs. 0.018) than in democratic country-years. The distribution for democracies is roughly normal, while for authoritarian countries it is right-skewed. In fact, most of the authoritarian samples in the WVS data do not show evidence of substantially inflated item nonresponse rates on the regime assessment questions. China is a notable exception, with consistently high scores on the self-censorship index across country samples. This suggests that self-censorship may be more prevalent in some authoritarian systems than others, which helps explain the divergence in findings in previous single-country studies (Frye et al. 2017; Jiang and Yang 2016; Wedeen 2015).

Thirdly, this variation across authoritarian countries is driven in part by political institutions (Boix and Svolik 2013; Gandhi 2008; Gandhi and Przeworski 2007; Svolik 2012). An exploratory analysis of the authoritarian samples shows that countries with electoral competition for the executive tend to have lower levels of self-censorship (Gandhi and Lust-Okar 2009; Levitsky and Way 2002; Levitsky and Way 2010; Morse 2012; Schedler 2013). Conversely, where the authoritarian ruler is chosen through a wholly non-competitive processes (internal appointment processes, hereditary succession, etc.), citizens are more reluctant to voice their political preferences. In these systems, the regime is truly the 'only game in town', and this dominance over society may breed political reticence.

The index proposed here can also be used to assess variation in self-censorship rates across groups within an authoritarian system. Probing deeper into the Chinese case, we see that self-censorship is likely more prevalent among older cohorts, which voice extremely high levels of support for the regime (Chen 2004; Dickson 2016; Tang 2016). Reticence is also higher among women, non-Party members, and citizens in the lower or middle class. These findings are replicated across the WVS and another recent nationally representative survey in China. Combined, these results suggest that self-censorship may be concentrated among citizens who are socially and politically marginalized. It also raises doubts as to the reliability of measurement of political attitudes in China (Chen 2004; Chen and Dickson 2008; Dickson 2016; Li 2004; Tang 2016).

This article offers the first assessment of self-censorship across time and space. Our proxy is imperfect, but we are able to show that respondents living in most authoritarian systems do not say 'Don't Know' to regime assessment survey questions at unusually high rates. If such respondents were really taking the survey in an environment of fear, we would expect more of them to refuse such questions entirely (Berinsky 1999; Berinsky 2002; Berinsky 2004; Berinsky 2008). The fact that they do not raises doubts about the actual sensitivity of these questions, and in turn whether feigning positive attitudes is as widespread as assumed.

Preference Falsification, Self-Censorship and Item Nonresponse

The concept of preference falsification is one of the most influential ideas in the study of authoritarian politics. Kuran's (1991) key innovation is that in authoritarian systems, there may be a disconnect between the level of support a citizen expresses for the regime in public (*public preference*) and her true level of support (*private preference*). At a given point in time, a citizen's private preference is effectively fixed, but her public preference is under her control. When a citizen who does not actually like the regime in private voices support in public, she is said to be engaging in *preference falsification* (Kuran 1991). This behavior is a subtype of the broader phenomenon of social desirability bias (Arnold and Feldman 1981), where people adjust their survey responses or public persona in order to conform to social norms.

We consider survey interviews to represent interactions in which citizens provide their public preferences, which may or may not align with their private preferences (which are unobservable to the researcher). To understand how preference falsification might manifest itself in surveys, consider the regime assessment question shown in Figure 1. This question elicits a respondent's general confidence in government and is commonly used in large-scale cross-country surveys.

In a given society, there is a true distribution of confidence in government among the population – the distribution of private preferences, to use Kuran's terminology. The figure depicts a hypothetical ten-person society, with the true population values (private preferences) shown on the left-hand side. There is a relatively low level of confidence in government in this case: the modal response is '1. None at all'. We also see that one citizen (Citizen 0) does not know her level of confidence in government. Some portion of 'Don't Know' answers in the population is normal, as not all citizens have well-formed political attitudes on every issue, and there are real cognitive costs to answering survey questions (Berinsky 2004; Berinsky 2008; Krosnick 1991; Tourangeau, Rips and Rasinski 2000).

The right panel depicts systematic preference falsification, as it might manifest itself in a survey like the WVS. Citizens in authoritarian societies may inflate their assessments of the government out of fear, leading to high expressed levels of confidence, satisfaction and so forth. Here, Citizen 2 has inflated her response from '3. Quite a lot' to '4. A great deal'. Citizens 4 and 6 have moved up to '3. Quite a lot' despite having little confidence in the government.

This inflation process fits Kuran's (1997) precise definition of preference falsification, but it is quite difficult to measure in practice. When we observe positive attitudes about authoritarian governments, it may because citizens genuinely like their governments, or it may be because they are inflating their responses.[2] We can never be sure if a citizen who says '3. Quite a lot' to an enumerator really feels that way.

The test proposed in this article relies on the measurement of item nonresponse (Berinsky 1999), a behavior that should offer clues as to whether or not questions are sensitive. In the right panel of Figure 1, Citizens 5 and 8 have chosen to voice a 'Don't Know' response (public preference), despite holding well-defined negative attitudes towards the regime (private preference). This is self-censorship, a behavior that is related to but distinct from the preference falsification Kuran (1997) describes. Self-censorship is more feasible to measure at scale, and it is conceptually important in its own right.

Note that item nonresponse can occur for two reasons. The first is when a citizen is unable to form a coherent assessment or answer a survey question – so-called nonattitudes (Converse 1964). This is Citizen 0 in the example from Figure 1, a person who has cognitive difficulty answering a question. The second occurs when there are social costs to expressing an opinion

---

[2]Indirect techniques, like list experiments (Blair and Imai 2012; Corstange 2009; Imai 2011), endorsement experiments (Blair et al. 2013; Bullock, Imai and Shapiro 2011) and randomized response (Gingerich 2010; Warner 1965; Zdep et al. 1979), seek to elicit truthful responses by providing the respondent some level of protection, but they have met some problems with implementation in practice. It is also difficult to implement these methods across multiple countries or to assess falsification historically.

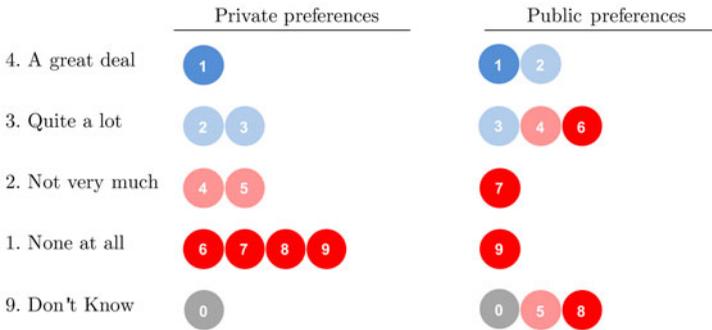Q1. How much confidence do you have in the government (in your nation's capital)?

Figure 1. Understanding preference falsification and self-censorship in surveys.

without sufficient personal or social benefits (Citizens 5 and 8). For example, in the contemporary United States, respondents might be unwilling to express attitudes that would be considered racist, so they might choose 'Don't Know' to such questions (Berinsky 1999; Kuran 1997). Thus in general, we observe higher nonresponse rates for questions that are cognitively complex, socially sensitive or both.

To summarize, we believe item nonresponse rates can be used to measure self-censorship, which will indirectly allow us to assess the broader idea of preference falsification, and to measure its likely incidence across time and space. If citizens in nondemocracies are really feigning positive attitudes for regimes they despise, we should see some indication in survey data that assessment questions are more sensitive in those contexts.

Data and Measurement Strategy

The article uses data from recent waves of the WVS. The WVS is a collaborative global survey project that measures citizen belief systems and their changes over time. To date, there have been six waves of the WVS, with almost 100 countries surveyed and over 400,000 interviews conducted. Many of the core political attitude questions of the WVS have been asked in multiple waves, which facilitates comparisons across time and space. Appendix Table A1 lists the authoritarian country samples used in the analysis.[3]

We calculate the following index across the different country-year samples:

$$\text{selfcens.ind}_{ct} = \left(\frac{\sum_{j=1}^{m} \overbrace{\sum_{i=1}^{n} \text{non response}_{ij}}^{\text{regime assessment}}}{m \times n} \right) - \left(\frac{\sum_{k=1}^{m} \overbrace{\sum_{i=1}^{n} \text{non response}_{ik}}^{\text{non sensitive}}}{m \times n} \right)$$

For each country c and sample at time t, for each regime assessment question j, we sum the item nonresponse indicator across all respondents n, sum this across all of the m questions, and divide the total by $m \times n$. This gives the mean item nonresponse for the regime assessment questions in the sample. We then subtract the mean item nonresponse for a set of nonsensitive political questions.

---

[3]Note that it is possible that citizens in authoritarian countries are less willing to take surveys, and this variation in response rates may affect the estimates shown here. The WVS research team does not report response rates for all of the samples, so we are unable to incorporate this into the analysis. We were able to track down response rates for a little over half of the samples in the WVS from the country reports (135/240), and it appears as though response rates in authoritarian samples are actually slightly higher – 77 per cent in authoritarian regimes vs. 68 per cent in democracies.

If we observe high levels of item nonresponse on sensitive questions in a given country-year sample, it may be because of self-censorship, or it may be because of low levels of political literacy and a lack of clearly defined attitudes (Berinsky 2004). Subtracting out nonresponse rates for non-sensitive questions corrects for the latter, and it gives us more confidence that the differential is capturing political self-censorship (Jiang and Yang 2016).

The questions used for the construction of the self-censorship index are shown in Table 1. The three regime assessment questions ask respondents to report their general confidence in government, and their perceptions of human rights and democracy in their country. Ex ante, these are the precise sorts of questions on which we would expect to observe self-censorship or preference falsification. The non-sensitive questions involve measures of interpersonal trust, life satisfaction and confidence in the television industry. All questions include a 'Don't Know' or 'No Answer' option, and some include both.[4]

Our goal was to pick questions with similar levels of cognitive difficulty but different levels of political sensitivity (Berinsky 2004). The choice of these questions is inherently arbitrary, and the index itself may be sensitive to this decision. We recommend other researchers explore different sets of questions in the index construction, and we have done so here. Our core index in the article relies on these six questions, but in the Appendix we assess the results across a two-question version of the index, a version that includes no nonsensitive questions, and a version that includes twenty nonsensitive questions. The core substantive findings of the article do not change appreciably.

Cross-National Results

Figure 2 shows the distribution of the index across regime type. Authoritarian and democratic country-years were identified using data from the Polity IV project (Marshall, Jaggers and Gurr 2002). To err on the side of inclusivity, we define authoritarian country-years as those having a combined Polity score of less than 5. The democratic countries represented in the figure correspond to 'full democracies' with a Polity score of 10. In order to expand the sample size, for some country-years, we use multiple imputation to impute item nonresponse rates for questions not included in the WVS (King et al. 2001). We also removed all country-year samples for which less than four of the six of index questions were administered.[5]

A few patterns emerge from this visual analysis. First, authoritarian country-year samples generally score higher on the self-censorship index (mean value of 0.022 vs. 0.014 for democracies). The coefficient estimate for a bivariate regression of the self-censorship index on democracy is statistically significant at conventional levels. This result is robust to comparing different versions of the index or using the full ordinal Polity measure of democracy (see Appendix Table A2 for estimates).

---

[4]The WVS research group appears to have left the decision about question wording for the item nonresponse categories up to the local research teams. Of the 240 country-year samples in the data, 210 appear to include a 'Don't Know' response for most questions, and 152 include a 'No Answer' response. About 53 per cent of samples (127) include both 'Don't Know' and 'No Answer', and the rest include only one of the options. In other uses of this type of index it might be possible to disaggregate 'Don't Know' and 'No Answer' responses, but that was infeasible for this project given the design of the WVS.

[5]Appendix Table A1 provides information on missingness. Our missingness model is at the country-sample level and includes the missingness rates for all six WVS questions (conf.gov, democracy, humanrights, conf.tv, gen.trust, life.sat) as well as an additional seventeen nonsensitive questions used to construct an alternative version of the self-censorship index. For the full data analysis, there are 1,194 total cells (199 country-year samples × 6 questions), each corresponding to a question-country-year. Of those cells, we are missing 158 (13.2 per cent). Of those 158 missing questions, 103 were missing because the WVS did not include the question at all. Another thirty-five were missing because the question was only administered in certain regions, and another four were omitted because of reasons related to questionnaire length. In the end, in only four of the 1,196 country-year cells does it appear that questions were possibly omitted for politically sensitive reasons: two in Uzbekistan (2011), one in Saudi Arabia (2003) and one in Qatar (2010). When we run tests on the nature of missingness proposed by Jamshidian and Jalal (2010), we fail to reject the assumption that the data is Missing Completely at Random. The number of imputed datasets is set to 50.

Table 1. Questions for falsification index construction world values survey – core questionnaire

Regime Assessment Questions	Non sensitive Questions
V115. I am going to name a number of organizations. For each one, could you tell me how much confidence you have in them: is it a great deal of confidence, quite a lot of confidence, not very much confidence or none at all? The government (in your nation's capital) (conf.gov)	V111. I am going to name a number of organizations. For each one, could you tell me how much confidence you have in them: is it a great deal of confidence, quite a lot of confidence, not very much confidence or none at all? Television (conf.tv)
V141. How democratically is this country being governed today? (democracy)	V24. Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people? (gen.trust)
V142. How much respect is there for individual human rights nowadays in this country? (human rights)	V23. All things considered, how satisfied are you with your life as a whole these days? (life.sat)

Secondly, the distribution of the index in the authoritarian country-year samples is right skewed, while for democracy it is roughly normal. The standard deviation of the index is also much lower for democracies (0.018 vs. 0.031). The figure shows that the distributions overlap quite substantially, and the difference in means is driven by only a few authoritarian cases. There are plenty of authoritarian samples in which citizens do not show inflated rates of item nonresponse.

The variation across authoritarian regimes is perhaps more interesting. Figure 3 shows the calculated self-censorship index and associated 95 per cent confidence intervals for the authoritarian country-year samples in the WVS. As a reference, the boxplots at the bottom of the figure compare the distribution of the index across regime types.

There are very clear examples of authoritarian countries where citizens seem reluctant to voice their assessments of the regime on surveys; China has among the highest values in the analysis for its three most recent WVS (2001, 2007, 2012) country samples. Morocco (2011, 2007), Pakistan (2001), Egypt (2001), Jordan (2007) and Azerbaijan (1997) also show some evidence of falsification.

However, self-censorship scores hover slightly above zero for most authoritarian countries, and within the normal range for most full democracies. In Singapore, Thailand, Kuwait, Mexico and Zimbabwe, there is little evidence that citizens self-censor on political questions. This is not to say that self-censorship does not exist in those countries; we simply do not find much evidence of it using our method.

It should be noted that some of the authoritarian regimes with the lowest Polity scores are excluded from this analysis. We do not have samples in the WVS for North Korea under Kim Jong Il, Syria under Bashar al-Assad or Cambodia under Pol Pot, for example. It is precisely those regimes where it is most difficult to collect public opinion data to begin with. It is possible that samples from these countries would show some self-censorship, with citizens saying 'Don't Know' on political questions. This may partly explain while we only find weak evidence of self-censorship in only a few systems: the harshest governments do not allow surveys like the WVS to be administered in the first place. This is an important caveat to the analysis.

What explains variation in the level of self-censorship in authoritarian systems? The Appendix reports the results of an exploratory data analysis that probes several plausible correlates of interest from the authoritarian politics literature: the level of repression (*repression*), the level of political competition in the executive branch (*exec.comp*), the presence of multiple parties (*party.comp*), the level of military control in politics (*military*) and the duration of the regime spell (*duration*). The repression variable is drawn from Wood and Gibney's (2010) Political Terror Scale (PTS), a five-point index where higher values correspond to higher levels of repression in society. All other variables are created from Svolik's (2012) Authoritarian Institutions dataset, which was extended by the authors to the present day to increase the coverage of the analysis.
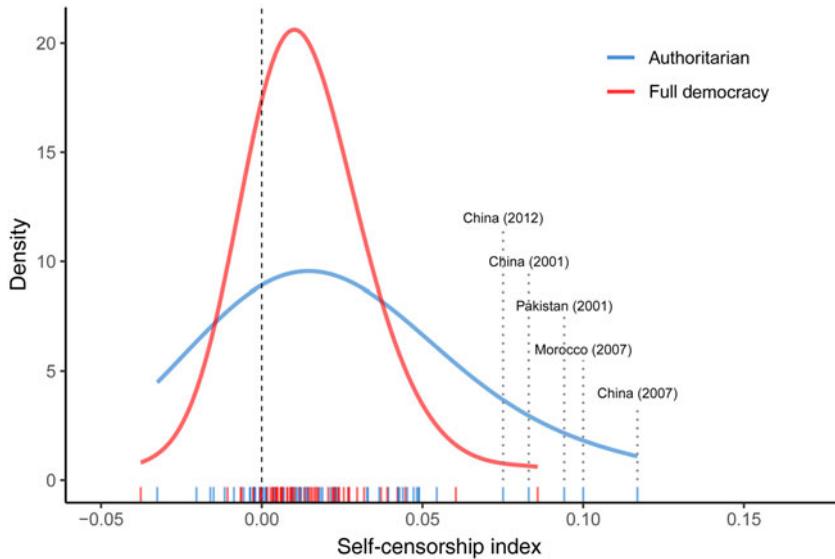
Figure 2. Self-censorship index by regime type.
Note: figure shows the distribution of the self-censorship index across democratic and authoritarian country-year samples. Democratic countries in the figure correspond to 'full democracies' with a Polity score of ten. Authoritarian samples are country-years with a Polity score of less than five.

The cross-national analysis produces two core findings. First, a lack of competition in the executive appears to be robustly associated with self-censorship. The *exec.comp* indicator is coded 1 for any authoritarian regime where the ruler is elected and 0 for all other regimes. This is meant to capture 'electoral authoritarian' regimes (Howard and Roessler 2006; Morse 2012; Schedler 2009; Schedler 2013). Paradigmatic cases in our dataset include Yeltsin's Russia, Mugabe's Zimbabwe and Zedillo's Mexico. Regimes with elections appear to have a roughly 2–4 percentage lower score on our index.

Though this analysis is strictly exploratory, we believe this finding reveals something larger theoretically about public opinion under authoritarianism. In systems without meaningful competition in the executive branch, the regime is truly the 'only game in town' and fully dominates the political system. In such closed authoritarian systems, opposition is weak or nonexistent (Miller 2013; Snyder 2006). This may facilitate cults of personality and ritualized demands of political loyalty (Wedeen 2015), where citizens may not be socialized into voicing opposition. It is intuitive that citizens are more likely to self-censor in these contexts.

The second finding of interest is that more repressive authoritarian regimes do not necessarily have higher levels of self-censorship. The measure of repression we use, the PTS, does not seem to be robustly associated with the falsification index (Wood and Gibney 2010), though there may be a weakly positive relationship. This result may be partially driven by the poor coverage of the WVS in hyper-repressive regimes, or the crudeness of the PTS itself. Either way, the analysis raises some doubts about the relationship between repression/fear and self-censorship, which have been echoed in recent experimental work (Lei and Lu 2017; Stockmann, Esarey and Zhang 2018).

Application: Self-Censorship in China

In the remainder of the article, we use the method proposed above to probe the nature of self-censorship in contemporary China. According to Howard and Roessler's (2006) typology, China under the CCP represents a closed authoritarian system. Chinese government leaders
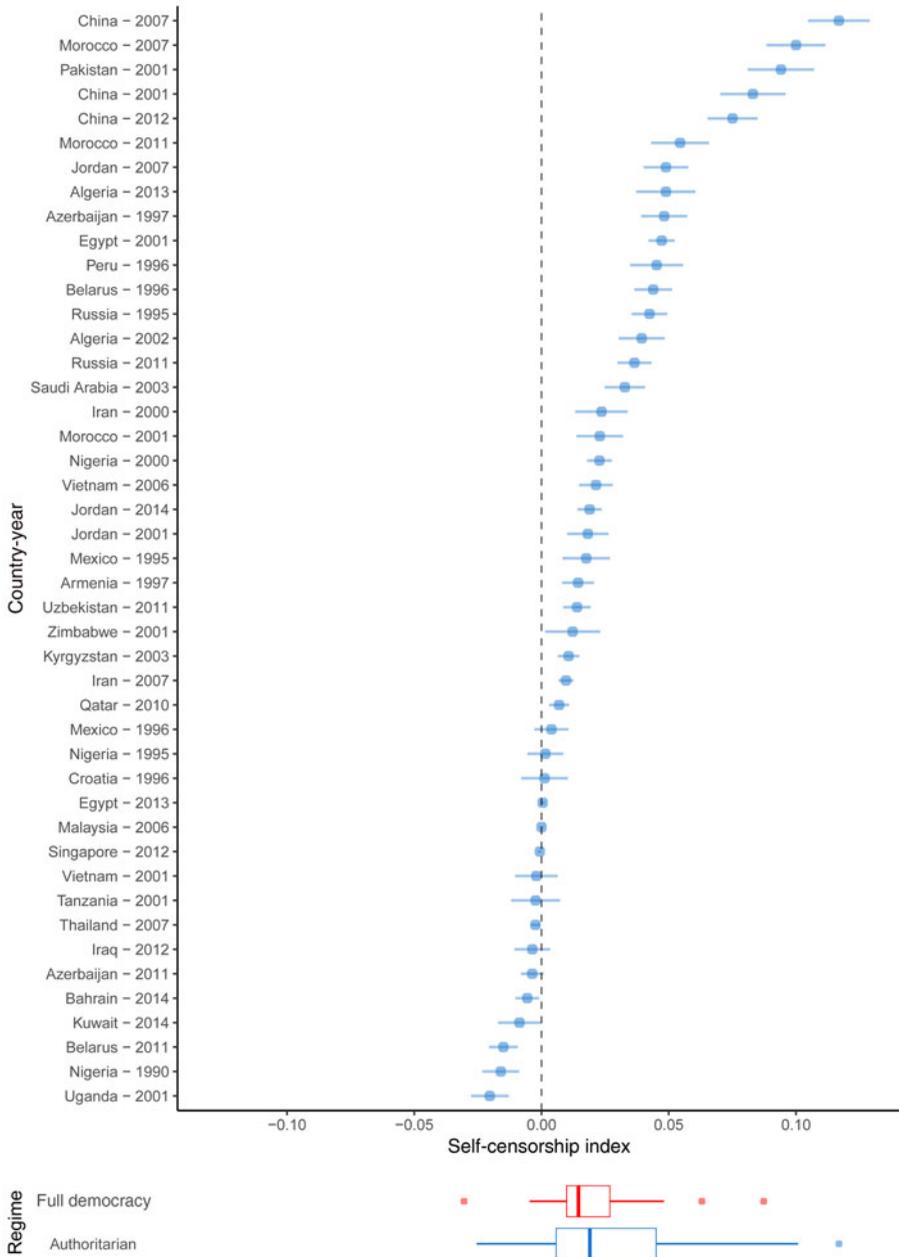
Figure 3. Self-censorship index in authoritarian country-year samples.
Note: figure shows the mean self-censorship index across different country-year samples of the WVS. Segments depict 95 per cent confidence intervals.

are chosen through an internal Party process, not a national election. This decision is ratified months later by the country's parliament (the National People's Congress) (Truex 2016). 'Democratic' parties exist in the parliament, but they are under the leadership of the CCP. These political institutions, combined with an increasingly sophisticated propaganda apparatus (King, Pan and Roberts 2013; King, Pan and Roberts 2017; Roberts 2018; Stockmann 2013), give the CCP near complete dominance of Chinese society.
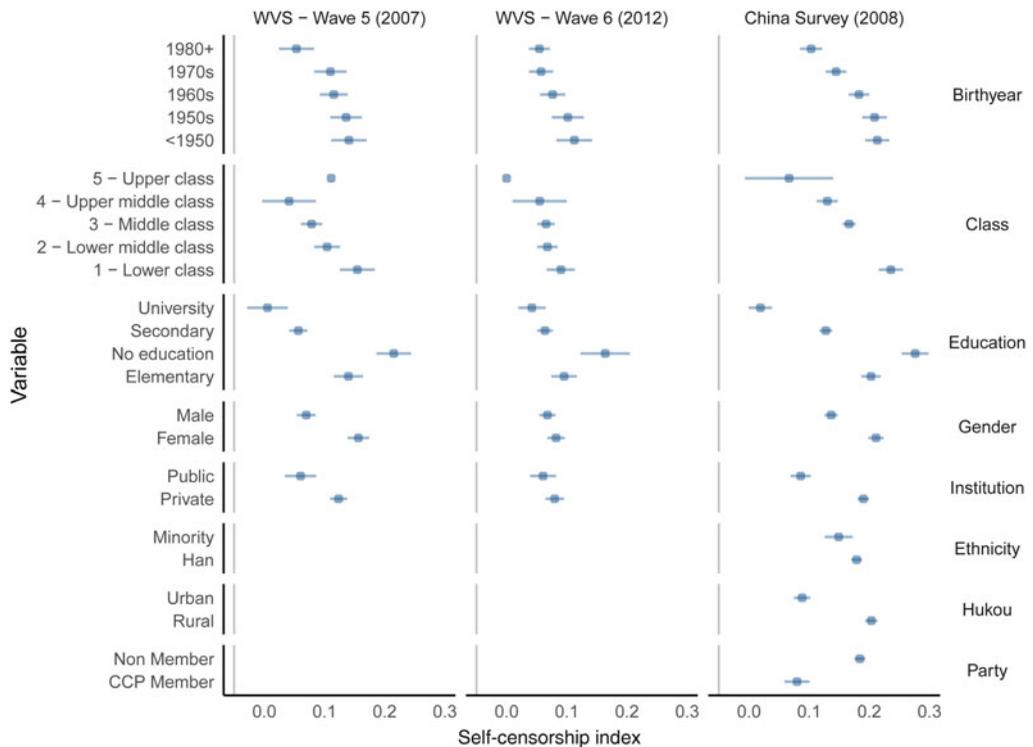
Figure 4. Self-censorship in China by subgroup.
Note: figure shows mean self-censorship index across different population subgroups in the China Survey and China samples of the WVS (Waves 5 and 6). Segments depict 95 per cent confidence intervals.

The analysis above raises doubts about the reliability of political attitudes reported by Chinese citizens. Chinese citizens do express positive attitudes about the government on surveys (Chen 2004; Chen and Dickson 2008; Dickson 2016; Li 2004; Tang 2005; Tang 2016; Truex 2017), but the analysis suggests that they refuse to answer these questions at unusually high rates. This is consistent with recent findings by Ratigan and Rabin (forthcoming), who find high rates of item nonresponse across a number of China surveys.

We can assess patterns across different subgroups within the population, provided the data are not too sparse. Figure 4 shows the self-censorship index for Chinese respondents in the WVS, aggregated by different demographic characteristics. The analysis is also replicated with additional groups for the 2008 China Survey, a nationally representative sample of over 3,000 respondents.

We find that self-censorship is higher among individuals who are marginalized in contemporary Chinese society: women, members of lower social classes, the less educated, citizens with rural household (*hukou*) registrations and non-Party members. These patterns hold largely constant across the three samples.[6] Older respondents in the sample are also more likely to self-censor on the sensitive political questions. For younger respondents – those born in the late 1980's and early 1990s – the 'Don't Know/No Answer' rates on sensitive questions are comparable to what we observe in advanced Western democracies. Younger citizens appear to be less socialized into self-censorship.

---

[6]Note that these findings contrast with those of Jiang and Yang (2016), who show that societal elites (high income, high education) were more likely to falsify their preferences following the political purge of Chen Liangyu in 2006, then the party secretary of Shanghai. This suggests that different groups may falsify in response to different political events.
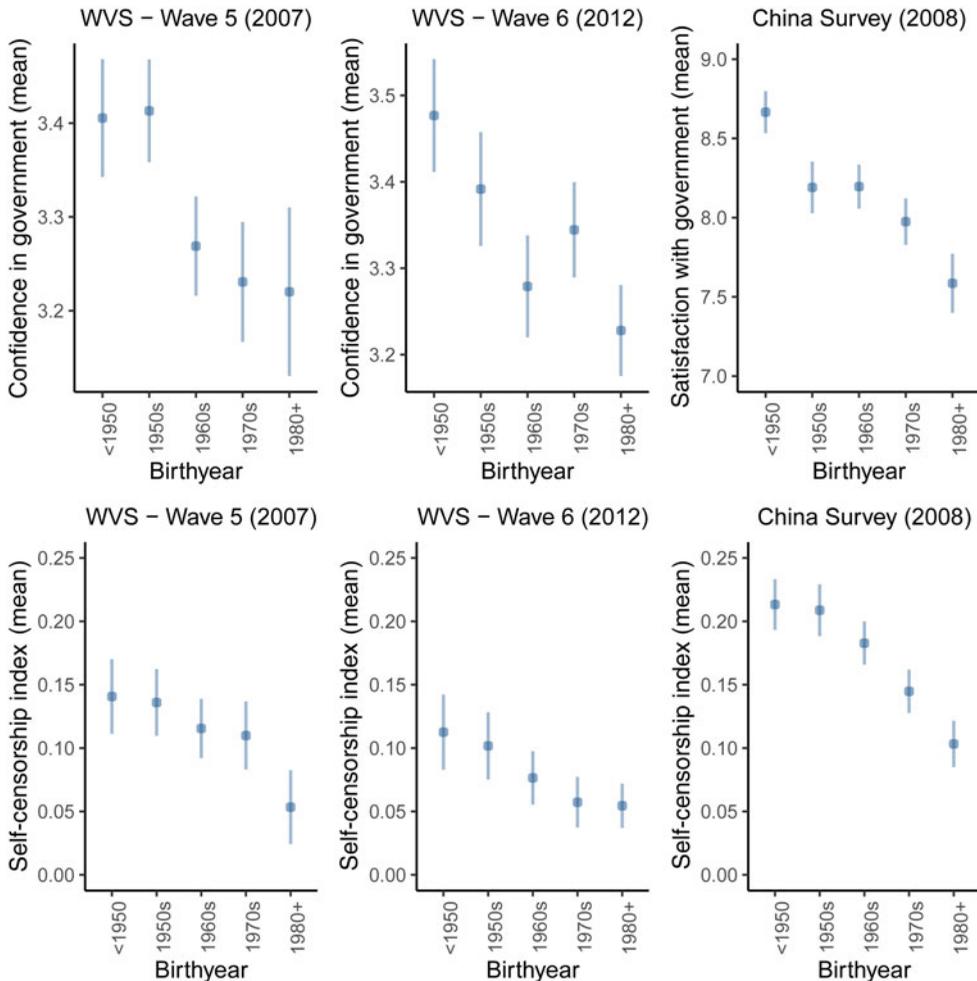
Figure 5. Regime support and self-censorship by age cohort (China).
Note: figure shows mean self-censorship index across different birth years for Chinese respondents in Waves 5 (2007) and 6 (2012) of the WVS and the 2008 China Survey. Segments depict 95 per cent confidence intervals.

For authoritarian societies with evidence of self-censorship, naïve descriptions of political attitudes within the population can be problematic (Chen and Dickson 2008; Dickson 2016; Li 2004; Tang 2005; Tang 2016). For instance, the top panel of Figure 5 shows the mean level of confidence/satisfaction in government for the WVS and China Survey, by birth year. A cursory look at the data would lead researchers to the conclusion that older respondents have more favorable views of the regime, as confidence in government declines with birth year (Chen 2004; Dickson 2016; Tang 2016). Yet we know that older Chinese citizens are also more likely to self-censor, so it is possible (and quite likely) that the elevated confidence levels are the result of less positive individuals opting out of the question entirely, or some inflation of responses.

Conclusion

For many authoritarian systems, citizens do not display higher rates of item nonresponse on direct regime assessment questions than their counterparts in democracies. This result suggests such

questions may not actually be all that sensitive in many places (Berinsky 2004), which in turn raises doubts that authoritarian citizens worldwide are widely feigning positive attitudes towards governments they secretly despise (Kuran 1991; Kuran 1997).

The data also show interesting heterogeneity in the authoritarian world, with self-censorship appearing limited in places like Thailand, Zimbabwe and Singapore, and countries like China and Morocco on the opposite end of the spectrum. Our analysis suggests that this variation is driven in part by the presence or absence of electoral competition in the executive branch. In electoral authoritarian regimes, the ruler faces at least some threat of losing office to an organized opposition, and this may give citizens more courage to voice their own opposition (Howard and Roessler 2006; Levitsky and Way 2002; Levitsky and Way 2010; Morse 2012; Schedler 2013). Political dynamics are different in China and other closed authoritarian regimes, where the absence of alternatives to the regime may foster the dynamics Kuran (1991, 1997) describes.

A wide array of tools is now available to assess sensitive political attitudes (Blair and Imai 2012; Blair et al. 2013; Bullock, Imai and Shapiro 2011; Corstange 2009; Gingerich 2010; Imai 2011; Warner 1965; Zdep et al. 1979), and the simple index proposed here is another low-cost method that can be scaled easily on pre-existing datasets. One promising avenue would be to probe variation in self-censorship over time: what types of events lead to shifts in item nonresponse rates? Repeated surveys have been implemented in many authoritarian countries (Treisman 2011), and the use of panel data could allow us to see whether repressive events or shifts in political competition affect citizens' willingness to voice their preferences.

Beyond mapping this variation, the next step in this research agenda should be to probe the psychology behind authoritarian rule. Our intuition is that preference falsification is unnatural, unpleasant and even painful, which is perhaps why we observe low levels of self-censorship in the authoritarian world. Citizens who truly despise the regime may choose to live authentically and voice their discontent, while others may rationalize their way into accepting authoritarianism. Both approaches seem more cognitively sustainable than 'living within the lie' (Havel 2009).

References

Arnold HJ and Feldman DC (1981) Social desirability response bias in self-report choice situations. *Academy of Management Journal* **24**(2), 377–385.

Berinsky AJ (1999) The two faces of public opinion. *American Journal of Political Science* **43**(4) 1209–1230.

Berinsky AJ (2002) Political context and the survey response: the dynamics of racial policy opinion. *Journal of Politics* **64**(2), 567–584.

Berinsky AJ (2004) *Silent Voices: Opinion Polls and Political Representation in America.* Princeton, NJ: Princeton University Press.

Berinsky AJ (2008) Survey non-response. In Donsbach D and Traugott MW (eds) *Handbook of Public Opinion Research.* Thousand Oaks, CA: Sage, pp. 309–321.

Berinsky AJ and Tucker JA (2006) 'Don't knows' and public opinion towards economic reform: evidence from Russia. *Communist and Post-Communist Studies* **39**(1), 73–99.

Blair G et al. (2013) Poverty and support for militant politics: evidence from Pakistan. *American Journal of Political Science* **57**(1), 30–48.

Blair G and Imai K (2012) Statistical analysis of list experiments. *Political Analysis* **20**(1), 47–77.

Boix C and Svolik MW (2013) The foundations of limited authoritarian government: institutions, commitment, and power-sharing in dictatorships. *The Journal of Politics* **75**(2), 300–316.

Bullock W, Imai K and Shapiro JN (2011) Statistical analysis of endorsement experiments: measuring support for militant groups in Pakistan. *Political Analysis* **19**(4), 363–384.

Chen J (2004) *Popular Political Support in Urban China*. Washington, DC: Woodrow Wilson Center Press.

Chen J and Dickson BJ (2008) Allies of the state: democratic support and regime support among China's private entrepreneurs. *The China Quarterly* **196**, 780–804.

Converse PE (1964) The nature of belief systems in mass publics. *Critical Review* **18**(1–3), 1–74.

Corstange D (2009) Sensitive questions, truthful answers? Modeling the list experiment with LISTIT. *Political Analysis* **17**(1), 45–63.

Dickson B (2016) *The Dictator's Dilemma: The Chinese Communist Party's Strategy for Survival*. Oxford University Press.

Frye T et al. (2017) Is Putin's popularity real? *Post-Soviet Affairs* **33**(1), 1–15.

Gandhi J (2008) *Political Institutions Under Dictatorship*. Cambridge: Cambridge University Press.

Gandhi J and Lust-Okar E (2009) Elections under authoritarianism. *Annual Review of Political Science* **12**, 403–422.

Gandhi J and Przeworski A (2007) Authoritarian institutions and the survival of autocrats. *Comparative Political Studies* **40**(11), 1279–1301.

Gingerich DW (2010) Understanding off-the-books politics: conducting inference on the determinants of sensitive behavior with randomized response surveys. *Political Analysis* **18**(3), 349–380.

Havel V (2009) *The Power of the Powerless: Citizens Against the State in Central-Eastern Europe*. London: Routledge.

Howard MM and Roessler PG (2006) Liberalizing electoral outcomes in competitive authoritarian regimes. *American Journal of Political Science* **50**(2), 365–381.

Imai K (2011) Multivariate regression analysis for the item count technique. *Journal of the American Statistical Association* **106**(494), 407–416.

Jiang J and Yang DL (2016) Lying or believing? Measuring preference falsification from a political purge in China. *Comparative Political Studies* **49**(5), 600–634.

King G et al. (2001) Analyzing incomplete political science data: an alternative algorithm for multiple imputation. *American Political Science Review* **95**(1), 49–69.

King G et al. (2004) Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review* **98**(1), 191–207.

King G, Pan J and Roberts ME (2013) How censorship in China allows government criticism but silences collective expression. *American Political Science Review* **107**(02), 326–343.

King G, Pan J and Roberts ME (2017) How the Chinese government fabricates social media posts for strategic distraction, not engaged argument. *American Political Science Review* **111**(3), 484–501.

Krosnick JA (1991) Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology* **5**(3), 213–236.

Kuran T (1991) Now out of never: the element of surprise in the East European revolution of 1989. *World Politics* **44**(1), 7–48.

Kuran T (1997) *Private Truths, Public Lies: The Social Consequences of Preference Falsification*. Cambridge, MA: Harvard University Press.

Lei X and Lu J (2017) Revisiting political wariness in China's public opinion surveys: experimental evidence on responses to politically sensitive questions. *Journal of Contemporary China* **26**(104), 213–232.

Levitsky S and Way L (2002) The rise of competitive authoritarianism. *Journal of Democracy* **13**(2), 51–65.

Levitsky S and Way LA (2010) *Competitive Authoritarianism: Hybrid Regimes after the Cold War*. Cambridge: Cambridge University Press.

Li L (2004) Political trust in rural China. *Modern China* **30**(2), 228–258.

Marshall MG, Jaggers K and Gurr TR (2002) *Polity IV project*. Center for International Development and Conflict Management, University of Maryland.

Miller MK (2013) Electoral authoritarianism and democracy: a formal model of regime transitions. *Journal of Theoretical Politics* **25**(2), 153–181.

Morse YL (2012) The era of electoral authoritarianism. *World Politics* **64**(1), 161–198.

Ratigan K and Rabin L (forthcoming) Reevaluating political trust: the impact of survey nonresponse in rural China. *The China Quarterly*.

Roberts ME (2018) *Censored: Distraction and Diversion Inside Chinas Great Firewall*. Princeton, NJ: Princeton University Press.

Schedler A (2009) Electoral authoritarianism. In Donsbach D and Traugott MW (eds) *Handbook of Public Opinion Research*. Thousand Oaks, CA: Sage, pp. 381–394.

Schedler A (2013) *The Politics of Uncertainty: Sustaining and Subverting Electoral Authoritarianism*. Oxford: Oxford University Press.

Shen X, Truex R (2019) "Replication Data for: In Search of Self-Censorship", https://doi.org/10.7910/DVN/OSSY4B, Harvard Dataverse, V1.

Snyder R (2006) Beyond electoral authoritarianism: the spectrum of nondemocratic regimes. In *Electoral Authoritarianism: The Dynamics of Unfree Competition*. Boulder, CO and London: Lynne Reinner, pp. 219–231.

Stockmann D (2013) *Media Commercialization and Authoritarian Rule in China*. Cambridge University Press.

Stockmann D, Esarey A and Zhang J (2018) Who is afraid of the Chinese state? Evidence calling into question political fear as an explanation for overreporting of political trust. *Political Psychology* **39**(5), 1105–1121.

Svolik MW (2012) *The Politics of Authoritarian Rule.* Cambridge: Cambridge University Press.

Tang W (2005) *Public Opinion and Political Change in China.* Palo Alto, CA: Stanford University Press.

Tang W (2016) *Populist Authoritarianism: Chinese Political Culture and Regime Sustainability.* Oxford: Oxford University Press.

Tourangeau R, Rips LJ and Rasinski K (2000) *The Psychology of Survey Response.* Cambridge: Cambridge University Press.

Treisman D (2011) Presidential popularity in a hybrid regime: Russia under Yeltsin and Putin. *American Journal of Political Science* **55**(3), 590–609.

Truex R (2016) *Making Autocracy Work: Representation and Responsiveness in Modern China.* Cambridge: Cambridge University Press.

Truex R (2017) Consultative authoritarianism and its limits. *Comparative Political Studies* **50**(3), 329–361.

Warner SL (1965) Randomized response: a survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association* **60**(309), 63–69.

Wedeen L (2015) *Ambiguities of Domination: Politics, Rhetoric, and Symbols in Contemporary Syria.* Chicago, IL: University of Chicago Press.

Wood RM and Gibney M (2010) The Political Terror Scale (PTS): a re-introduction and a comparison to CIRI. *Human Rights Quarterly* **32**, 367.

Zdep SM et al. (1979) The validity of the randomized response technique. *The Public Opinion Quarterly* **43**(4), 544–549.