Lying or Believing? Measuring Preference Falsification from a Political Purge in China[1]

Junyan Jiang[2] Dali L. Yang[3]

This Version: 05/17/2015

Abstract

Despite its wide usage in explaining some nontrivial regime dynamics in nondemocracies, preference falsification remains an empirical myth for students of authoritarian politics. We offer the first quantitative study of this phenomenon in an authoritarian setting using a rare coincidence between a major political purge in Shanghai, China, and the administration of a nationwide survey in 2006. We construct two synthetic measures for expressed and actual political support from a set of survey questions and track their changes over time. We find that after the purge there was a dramatic increase in expressed support among Shanghai respondents, yet the increase was paralleled by an equally evident decline in actual support. We interpret this divergence as evidence for the presence of preference falsification. We further show that falsification was most intense among groups that have access to alternative information but are vulnerable to state sanctions, and use two additional surveys to explore the inter-temporal dynamics of falsifying behaviors.

[1] Earlier versions of this article were presented at the annual meetings of the Midwest Political Science Association and American Political Science Association, as well as the Harris Political Economy Lunch, the Comparative Politics Workshop and the East Asia Workshop at the University of Chicago. We thank Michael Albertus, Milena Ang, Sofia Fenner, Anthony Fowler, Yue Hou, Stanislav Markus, Andrew Mertha, Pablo Montagnes, Monika Nalepa, Lisa Wedeen, Stephane Wolton, Yan Xu, Yang Zhang and participants at aforementioned workshops for providing helpful comments and feedback. We also thank Shizheng Feng for generously sharing the CGSS data, Zhuang Liu for his help on the online experiment, and Yalin Liu for excellent research assistance.

[2]Graduate student, Department of Political Science, University of Chicago

[3] Professor, Department of Political Science, University of Chicago

I. Introduction

The human tendency to deliberately misrepresent information about themselves has been found in numerous settings and studied by various social science disciplines. This tendency is particularly strong in the realm of politics, where individual actions bear public consequences. A large body of research on democracies has found that people hide negative feelings toward certain race, gender or religious groups (Gilens, Sniderman, & Kuklinski, 1998; Kane, Craig, & Wald, 2004; Kuklinski, Cobb, & Gilens, 1997; Kuklinski, Sniderman, et al., 1997; Powell, 2013; Streb, Burrell, Frederick, & Genovese, 2008), lie about having voted in elections (Bernstein, Chadha, & Montjoy, 2001; Holbrook & Krosnick, 2010; Karp & Brockington, 2005; Silver, Anderson, & Abramson, 1986), and underreport desirable activities such as involvement in corruption and vote buying (Corstange, 2010; Gonzalez-Ocantos, de Jonge, Meléndez, Osorio, & Nickerson, 2012).

If lying about political attitudes and behaviors is common in democracies, it is probably essential to survival in autocracies. Lacking legitimacy from a popular mandate, authoritarian regimes routinely sanction those who dare to openly challenge the dominant discourse that is deployed to sustain the regime (Arendt, 1972; Linz, 2000). To avoid imminent punishment from the state, ordinary citizens under those systems often feign support for things that they privately oppose (Havel, 1990). This act, best known and analyzed as preference falsification by Kuran (1991, 1997), is believed to be consequential to the persistence of authoritarian regimes (Kuran, 2011; Kurzman, 2009; Patel, 2012; Wedeen, 1999).

Yet, despite its wide usage in studies of nondemocratic systems, the concept of preference falsification remains in many ways an empirical myth for students of authoritarian politics. What

is the scope of preference falsification in authoritarian regimes? Under what conditions and on which issues are citizens likely to falsify their preferences? Is the tendency to falsify stronger for certain social groups than others? To date, even the first question has not been adequately answered. Although many influential theoretical propositions hinge on the assumption that preference falsification has a constant and pervasive presence in nondemocratic systems, this assumption is rarely tested against systematic data. The strongest evidence supporting this assumption so far comes from personal accounts and selected ethnographic case studies (Havel, 1990; Wedeen, 1999), which cannot be easily generalized to the population at large, or to all forms of autocracies. Others, however, have also found that some authoritarian regimes, such as post-reform China and Mubarak's Egypt, have a fairly open public sphere with a diverse set of opinions that seem, at least on the surface, free of severe falsification (Blaydes, 2010; King, Pan, & Roberts, 2013; Tang, 2005).

Nor do studies based on survey data provide any definite answers to these questions. Granted that early researchers working with surveys from communist regimes did find some inconsistencies in responses that were indicative of preference falsification (Sulek, 1994; Welsh, 1981), the bulk of recent works on public opinions in authoritarian regimes have mostly avoided a direct discussion of this issue, either by proposing a better research design that seeks to achieve as-if random assignment on unobservable factors (including the propensity to falsify therefore) (Kern & Hainmueller, 2009; Lü, 2014), or by simply ignoring such biases altogether. In light of the challenges that empirical researchers face in obtaining reliable public opinion data from nondemocratic systems, the dearth of empirical research on this topic is understandable but it would be especially valuable if we can gain a better understanding of the dynamics of preference falsification in authoritarian systems. Most importantly, it would help researchers to more

accurately assess the actual strength of an authoritarian regime and the prospects for change: If a regime's popularity is maintained by force and intimidation, then it is fragile despite the appearance of unity and it may be highly unstable if information about dissatisfaction becomes widespread (Kuran, 1991, 2011; Kurzman, 2009; Lohmann, 1994). In contrast, an authoritarian regime founded on genuine public support is more robust and likely to withstand even major political turbulences and shocks. Moreover, if future political changes follow a process of revolutionary bandwagon as suggested by Kuran (1991), then knowing the distributions of genuine and feigned supporters in a society would help us better forecast how such changes might unfold.

In this article, we provide to our knowledge the first systematic study of the phenomenon of preference falsification in an authoritarian system by examining public responses to a major political purge in China. On September 25, 2006, Chen Liangyu, then the Party Secretary of Shanghai and a member of the Politburo of the Chinese Communist Party Central Committee, was abruptly dismissed from his posts and placed under investigation for corruption. Around the same time, the Chinese General Social Survey (CGSS), a nationwide survey sponsored by a major Chinese university, was being conducted across China and Shanghai was included in the sampling frame. The occurrence of the purge naturally divided the sample into two groups, a "treatment" group that was interviewed after the purge and a "control" group that was interviewed before that. While in no way designed for our study, the survey nevertheless included two types of questions relevant for our research purpose: (1) a set of sensitive questions that explicitly ask respondents' *political support* for the state, and (2) another set of questions that are less sensitive but nonetheless *reflective of respondents' true attitude* towards Chen's dismissal. We construct two synthetic measures to capture the *expressed* and *actual support* for

the purge from a number of survey items, and track their changes before and after the purge. Assuming that the progression of the survey is largely uncorrelated with respondents' characteristics—a point we will demonstrate later—this quasi-experimental setting provides a unique opportunity for assessing how respondents in Shanghai reacted to a major political event in terms of both their expressed attitudes and underlying beliefs.

We find that in the aftermath of the purge the average Shanghai became much more likely to give affirmative answers to sensitive survey questions explicitly linked to political support. At the same time, however, they were much less happy, more cynical about the legal system but remained reluctant to update their beliefs about the state of official corruption, suggesting that they have little genuine trust in the official anti-corruption narrative. To us, the divergence between expressed and actual support is indicative of the presence of preference falsification in the immediate aftermath of the purge. We show that our findings are robust to different methods in constructing the synthetic measures, and perform several tests to demonstrate that the estimated effects are unlikely to be caused by other concurrent events.

We also analyze how the degree of preference falsification varied across subgroups. We conjecture that preference falsification would arise under the confluence of two factors: access to alternative sources of information and perceived vulnerability to state sanctions. The variations we found are largely consistent with the conjecture: preference falsification is most intense among the wealthy, the highly educated, frequent internet users and state-sector employees.

Finally, we examine how preference falsification changed over time after the purge by comparing responses to similarly worded questions in two other nationwide surveys conducted one year apart from each other. We show a sharp decline in reported trust in both the central

government and the Chinese Communist Party (CCP) among Shanghai residents over this period. This finding further suggests that effective political persuasion did not happen even in the long run. However, people did become less concerned about expressing their negative feelings under a much relaxed political environment.

Our study demonstrates the systemic presence of preference falsification in an authoritarian regime and reveals its temporal dynamics. It contributes to several strands of research. First and foremost, we contribute to the understanding of mass political behavior in nondemocratic settings. Contrary to the notion that preference falsification is no longer a serious issue for studies of authoritarian states in the age of information, we show that it still exists on a large scale even in a substantially open system like China's. When it comes to important political issues on which the state has a clear position, ordinary citizens still do not dare to publically reveal their disagreement with the dominant discourse. Meanwhile, our study updates the classical perspectives on authoritarian politics by showing that the public's propensity to falsify their preferences is better construed as a *variable* than a constant. The propensity of the citizens to falsify their attitudes is not only a function of the larger political environment but also has important individual-level determinants.

Second, our study sheds light on the limits and utility of social surveys in informing researchers about public opinion in non-democracies. Experts in comparative politics are sharply divided on this issue: while some are skeptical about the reliability of such social survey data (Darden & Grzymala-Busse, 2006; Rose, 2007), others see considerable value in them for understanding authoritarian politics (Manion, 2010; Tang, 2005; Tessler, 2002). While we recognize surveys in authoritarian regimes face formidable problems, we argue the biases can be detected and overcome by using carefully chosen questions and by incorporating deep contextual

knowledge about survey timing into analysis. While more powerful experimental techniques are still being introduced to the study of authoritarian politics (Malesky, Schuler, & Tran, 2012; Meng, Pan, & Yang, 2014), the analytical strategies we take in this article provide useful examples for obtaining more reliable inference from surveys that already exist.

By assessing the impact of the purge on overt and actual political support and tracking their changes over an extended period of time, our study also contributes to a growing literature on the nature and sources of political support in authoritarian systems, and in China in particular (Chen, 2004; Geddes & Zaller, 1989; Lewis-Beck, Tang, & Martini, 2013; Rose, Mishler, & Munro, 2011; Treisman, 2011).While existing studies emphasize the role of economic performance in sustaining regime support (Zhao, 2009), we highlight that, at least in the Chinese context, moral and political performance are equally salient, if not in fact more, considerations in people's assessment of regime legitimacy (Zhao, 2009). Our analysis of Chen Liangyu's dismissal demonstrates that a corruption investigation that was widely perceived to be disingenuous and politically motivated could have deleterious effects on regime legitimacy. The purge of Chen not only produced intense emotional distress and mistrust among the local population in the short run, but it also led to substantial decline in general political support that persisted in the medium to long run.

The rest of the article is structured as follows. We begin by introducing the context of the purge and the national survey that we use. In the third section we propose our measurement strategy and discuss its construct validity. Sections 4 through 7 present the estimation framework and the results. We discuss the implications of our findings in section 8 and conclude.

II. Background

The Purge

Our analysis of preference falsification was made possible by the controversial dismissal of a high-ranking official in China. On September 24, 2006, Chen Liangyu, then Shanghai's Party chief and a sitting member of the Politburo, was abruptly dismissed and placed under disciplinary investigation. In the official announcement, which was broadcast nationwide on national TV the following evening, Chen was accused of multiple corrupt dealings, including misuse of Shanghai's social security fund for the benefits of his business friends, shielding colleagues who committed serious crimes, and abusing his position to secure benefits for family members.[4]

Chen's purge was extensively covered by both domestic and overseas media. Chinese media toed the official line and condemned Chen's corrupt behavior.[5] In contrast, most overseas media viewed the purge of indicative of a power-struggle between Hu Jintao, then the incumbent Party General Secretary and President, and the Shanghai-based faction headed by former General Secretary and President Jiang Zemin, with whom Chen was closely associated. On the latter account, Chen was sacked for his disrespect to the new leadership and stubborn resistance to the center's macroeconomic sterilization policies at a time when Hu and Premier Wen Jiabao were still seeking to establish their authority (Fong, 2004; Li, 2007)

While we do not know the exact share of people who subscribed to each narrative, both likely had some believers in the post-purge Shanghai. In particular, even though the power-struggle narrative was heavily censored in Chinese media, it might still have attracted a fairly large

[4] "Decision to Investigate into Comrade Chen Liangyu's Serious Disciplinary Violations." 2006. *Xinhua News Agency*.

[5] In an analysis of 50 articles on Chen's dismissal published in the domestic media, we find that the word "anti-corruption" (反腐败) alone appeared 80 times, followed by "serious violations" (严重违纪) and "law and discipline" (党纪国法), which occurred 48 and 33 times, respectively.

audience among the local residents. Before his dismissal, Chen had enjoyed a quite high level of popularity in Shanghai, both due to his effectiveness as an administrator and his background as a local native. In a survey of popular approval for mayors in10 large Chinese cities, for example, Chen, who was then the mayor of Shanghai, received the highest level of approval from people in his jurisdiction among all mayors surveyed.[6]

In Figure 1, we plot the temporal and spatial distributions of internet search activities (measured by Google search index) for three key words (Chen Liangyu, Corruption and Shanghai Gang) associated with different underlying interpretations of the event.[7] For all three keywords, the search intensity peaked in the week immediately following the purge. Most notably, the keyword "Shanghai Gang" received as much as 30% of the search intensity for the word "Corruption" despite the heavy censorship on the former. The bottom panel of Figure 1 further shows public attention to this event concentrated in Shanghai, with the aggregate search intensity over three times higher than the second highest region (Jiangsu).

[Figure 1 about here]

The Survey

The main data used in this study come from the China General Social Survey (CGSS), China's longest-running professional national survey based at Renmin University of China since 2003. The 2006 wave of CGSS was conducted between September and November and interviewed 10151 individuals from 28 provincial units. The CGSS sample in Shanghai in

[6] The survey is conducted by the Horizon Group, a private survey company in China.

[7] We choose these three keywords because they are related to difference underlying interpretations of the event: neutral (Chen Liangyu), anticorruption (Corruption) and power-struggle (Shanghai Gang). We only focus on search from mainland China between April 2006 and April 2007. Google was then one of the two top search engines within China.

included 400 individuals in five major urban districts and the interviews were carried out between September 11 and November 12.[8] Chen's dismissal was first announced on national TV at 7:00 pm on September 25, by which all the CGSS interviews on that day had been completed. As a result, we thus use September 26 as the start of the treatment period. For comparability, we focus on sample provinces with *on-going* surveys on September 26 and exclude sample provinces where *all* respondents were interviewed either before or after that date. This reduces our sample to 5,046 observations in 12 provinces, with 2,638 respondents (281 in Shanghai) in the treatment group and 2363 (119 in Shanghai) in the control. A detailed description of the survey is in the Appendix.

III. Measuring Preference Falsification: Explicit and Implicit Measures

Selecting Questions

To assess public responses to Chen's dismissal and, more importantly, to distinguish preference falsification from genuine conviction, would require us to obtain information about both respondents' expressed and actual attitudes toward the purge. The challenge of achieving this in a conventional survey is twofold: First, we need to separate the effect of a given event on respondents' attitudes from other confounding factors in order to say something about people's opinions about a specific event of interest. Second, in addition to expressed attitudes, we also need to measure respondents' underlying beliefs in order to differentiate untruthful responses from truthful ones.

Although CGSS did not directly ask respondents' attitudes toward Chen's dismissal, the natural coincidence between the announcement of the dismissal and the survey allows us to

[8] The five sampled districts are Xuhui, Yangpu, Jingan, Putuo and Pudong.

overcome the first challenge by comparing responses to an identical set of questions before and after the purge. If nothing significant happened during this period, the responses from the treatment and the control periods should be roughly comparable. Any systematic difference between these two groups, on the other hand, can be attributed to events that occurred only in the treatment period. As for the second challenge, we assess the respondents' sincerity by comparing their responses to two different sets of questions: a group of questions that explicitly ask respondents' political loyalty and support, and another group of questions that are less politically sensitive but indirectly related into respondents' attitudes toward Chen's dismissal. Our approach shares the same spirit with the use of unobtrusive measures in social psychological research (Fazio & Olson, 2003; Webb, Campbell, Schwartz, & Scechrest, 1966). The basic idea here is that while respondents may intentionally manipulate their answers to questions that have an explicit goal of measurement and a clear, socially desirable answer, they are less motivated to (and probably also less able to) lie when the purpose of the question is vague and when there is not a single "right" answer. If respondents are sincere, we would expect their answers to the implicit measures to be consistent with their answers to the explicit ones. Inconsistency between their responses to the two sets of measures, on the other hand, is evidence for preference falsification. We formalize this idea with a simple model in the Online Appendix.

In the specific context of our study, we choose the following questions as our explicit measures for attitudes

- **[Compliance]** *"Do you agree or disagree with the following statement: Following the government can never be wrong"*

  (1. Strongly disagree 2.Disagree 3.Agree 4.Strongly agree.)

- **[Development over Democracy]** *"Do you agree or disagree with the following statement: There is no need to raise the level of democracy as long as the economy keeps growing."*

  (1. Strongly disagree 2.Disagree 3.Agree 4.Strongly agree.)

- **[Government over Law]** *"Do you agree or disagree with the following statement: We must respect government's decisions over those of the court when the former contradict the latter "*

  (1. Strongly disagree 2.Disagree 3.Agree 4.Strongly agree.)

Clearly, all three questions measure some aspects of general political support.[9] The first and the third questions ask explicitly about respondents' trust in and respect for the political authority, and the second one is closely related to attitudes toward the current regime, with a hint to democratization as the alternative. If there was no political pressure, one would expect those who believed that Chen was indeed corrupt to be more likely to support the Center's action and answer those questions in a more affirmative tone in the immediate aftermath of the purge. In contrast, those who sympathized with Chen would likely view the move negatively and be less likely to agree with those statements. However, to the extent that answers to these questions are directly related to one's political loyalty, it creates incentives for disgruntled respondents to hide their negative feelings, especially when there was explicit pressure for displaying political support. If this was the case, then respondents would give more affirmative responses to these

[9] Although the questions are framed in terms of attitudes toward the government rather than the Party, the majority of the Chinese respondents do not make strong distinction between these two entities. So these questions can be understood as demonstrating attitudes toward the political authority in general.

questions *regardless* of their underlying beliefs, and the observed level of support from these questions would be biased upward from the actual level of support.

For implicit measures, we pick the following questions:

- **[Happiness]**"*Overall how you would describe your life?*

  (1. Very unhappy 2. Unhappy 3. So so 4. Happy 5. Very unhappy)

- **[Judicial Independence]**"*Do you agree or disagree with the following statement? The court always agrees with the government on major cases."*

  (1. Strongly disagree 2.Disagree 3.Agree 4.Strongly agree.)

- **[Beneficiary Group]** *"Which of the following groups of people in your opinion have benefited most in the last ten years?"*

  (1. Peasants 2. Workers 3.Government officials 4. Foreign investors…)

We reverse the scale for the second question (making "strongly disagree" to 4 and "strongly agree" to 1) to be consistent with the labelling. For the last question, we recode the responses into a binary variable which takes the value of 1 if the respondent chooses "government officials" as their answer and 0 otherwise.

Compared to the explicit measures, the implicit measures do not appear to directly pertinent to one's attitudes toward the regime as a whole. The second and third questions ask for assessments of certain parts of the system (courts and officials) only and are not about the regime as a whole (in follow-up experiments, we will show that there is no consensus on which answers are more politically desirable). In the specific political context of Chen's dismissal, we have good reasons

to believe that respondents would answer these questions in systematically different ways depending on their beliefs about the nature of Chen's dismissal. Those who bought in and supported the Center's anti-corruption narrative would likely be happier (because a major corrupt official was uncovered and being punished), more likely to choose "government officials" as the greatest beneficiaries in the past decade (by updating their beliefs from the corruption scandal), but have their views about the judiciary largely unchanged.

In contrast, those who sympathized with Chen and were skeptical about the official narrative would be expected to show exactly the opposite attitudinal responses. Their happiness measure would decline as they saw the dismissal of the capable Chen as driven by the central leaders' political machinations. Since the Party's disciplinary investigations would inevitably be followed by a guilty conviction (Manion, 2004), Chen's purge would heighten their sense of the lack of judicial independence when they were asked about the legal system. Moreover, as they saw the anticorruption case against Chen as illegitimate, their response to the third question would stay largely unchanged.

The predicted changes for supporters and opponents of the purge are summarized in Table 1. We also test the validity of our predictions in a follow-up online experiment where we inform the subjects of a dismissal of a high-ranking official using one of two one-sided frames (either on anticorruption or power struggle) and then ask them to respond to the original survey measures. The results of the experiment are largely consistent with our predictions. The details of the experiment can be found in the Online Appendix.

[Table 1 about here]

Testing Differential Sensitivity

The validity of our approach depends on the critical assumption that our explicit and implicit measures of attitudes are *differentially sensitive* to respondents. We provide several tests to verify this assumption. To begin with, we conducted a survey experiment (n=311) on one of China's leading online commercial survey networks (www.sojump.com) in July 2013 and asked two groups of respondents to assess the perceived sensitivity of a list of questions, including both our explicit and implicit measures.[10] Following the method developed by Bradburn, Sudman, and Blair (1979), we ask one group to give answers that *most people would give*, and the other to give *their real answer*. If a question is politically sensitive, we should see a significant difference in the answers given by these two groups. (A detailed description of the experimental protocols is available in the Online Appendix.) We display the results in the first two columns of Figure 2. In the first column on the left, we see that all three explicit measures are rated as more sensitive than the implicit measures. Similarly, in the second panel where we plot the average difference in responses between the "most people's answer" group and the "your real answer" group, there were significant differences between the two groups for all explicit measures but no such difference for any of the implicit measures.

[Figure 2 about here]

In the third column of Figure 2, we further gauge the relative sensitivity of the questions by looking at the missing rate from the actual data. Intuitively, the incidence of missing responses will be higher for more sensitive questions. Here we exclude the question of **[Beneficiary Group]** as the original question did not provide a "don't know" or "no-response" option. For the other five questions, the pattern is consistent with the experimental results: the percentages of

[10] We randomize the order by which these questions appear to eliminate any anchoring effects. See the Appendix for a more detailed discussion of the experiment protocol.

missing response are on average higher for the explicit measures than implicit ones. Within the Shanghai sample, in particular, the explicit measures have an average missing rate of 9.67%, whereas the question on **[Judicial Independence]** only has a missing rate of 5.8% and there is no missing values for the question on **[Happiness]**.

IV. Empirical Strategies

We use a difference-in-difference (DID) approach to estimate the effects of Chen's dismissal on Shanghai respondents in terms of both expressed and actual support. Specifically, we estimate the following equation:

$$\text{Attitude}_{ijt} = \delta \text{Treatment}_{ijt} \times \text{Shanghai}_{ijt} + \mathbf{X_{ijt}\beta} + \eta_j + \theta_t + \epsilon_{ijt},$$

where i indexes the respondent, j the district and t the date of interview. For the dependent variables, we not only use the original measures, but also create two synthetic indices, which we hereafter refer to as *Expressed support* and *Actual support*, from the original measures in order to reduce noise and simplify interpretation. We use both simple averaging and factor analysis to construct these synthetic indices (see Online Appendix for details). *Treatment* is a binary variable that takes the value of 1 if the respondent was interviewed on or after September 26 2006 and 0 otherwise and *Shanghai* is a dummy for the region of interest. The key quantity of interest, δ, is the average treatment effect (ATE) of Chen's dismissal on the attitudes of respondents in Shanghai. $\mathbf{X}$ is a vector of demographic covariates. η denotes the district-level fixed-effect and θ the date fixed effects. While we do not explicitly write them in the equation, the main effects for the *Treatment* and the *Shanghai* dummies are subsumed under the district and date fixed-effects in this specification.

In addition to the standard parallel trend assumption, which we verify in the next section, our specific research design requires "as-if" random assignment of the "treatment" and "control" status, so that those interviewed in these two periods have comparable counterfactuals. Although there is little evidence that the order of interviews was in any way systematically correlated with respondent characteristics,[11] non-strategic selection in treatment assignment might still arise due to logistic arrangements of the survey, or even by chance given the sheer size of the sample. We provide a set of balance tests in the Online Appendix to show that the differences on observable attributes between respondents in the treatment and control groups are substantively small, and, at least at the aggregate level, there is little serial correlation in respondents' characteristics across different survey dates. To further address the existing imbalance, we use entropy balancing (EB), a re-weighting technique developed by Hainmueller (2011), to adjust the distributions of covariate in treatment and control groups. The EB algorithm calculates a weighting scheme that achieves almost exact balance on the first and higher movements of targeted covariates specified by the researcher. We choose to balance on a set of key demographic variables, including *Age*, *Gender*, *Religion*, *Occupation*, *Education*, *Marital status*, *CCP membership*, *Military experience*, *Employment status*, and *Urban residence*. The difference in these observable characteristics between the treatment and the control became virtually 0 in the reweighted data.

V. Results

[11] To test this, we aggregate a number of observable demographic by survey date and estimate the autocorrelations between day *t* and day *t-k*, and we detect virtually no correlation for any value of *k* between 1 and 15.

In this section we present the main results on how respondents in Shanghai responded to Chen's dismissal in terms of both expressed and actual attitudes. We begin by visually examining how responses changed over time: In Figure 3, we plot the changes in two synthetic measures based on the first principal component and their difference (all standardized) over the duration of the survey. On each panel, we fit four separate 5^{th} order polynomial lines for all the combinations of treatment status (treatment vs. control) and region (Shanghai vs. the rest of the country). Reassuringly, the parallel trends assumption looks largely valid: For both synthetic measures, the trajectories were almost identical between Shanghai and the rest of the country during the period leading up to the announcement of Chen's dismissal, but started to diverge immediately afterwards. Expressed support in Shanghai went significantly above national trend whereas actual support in the city dropped significantly below.

[Figure 3 about here]

Next, we use regression analyses to assess the significance of these visual patterns. Table 2 presents the regression estimates for the effects of the purge on responses to explicit and implicit measures for Shanghai respondents as a whole. In each panel, we first present the estimates using the original measures as the dependent variables. The last four columns present the estimates using the synthetic measures created by both simple averaging and the principal component methods. For each of the synthetic measures, we run two models, one with our full sample and the other with a smaller sample that only includes respondents interviewed prior to November 1^{st}. The exclusion of the post-Nov 1 interviews is motivated by the consideration that the downfall of Chen might no longer be salient to respondents more than a month after its occurrence.

[Table 2 about here]

The regression results are consistent with our impressions from the visual analysis: The dismissal of Chen on corruption grounds had substantial impacts on both the expressed and the actual attitudes of Shanghai residents', but in *opposite* directions. For the explicit measures, the coefficient estimates are positive and significant for all but one original measure as well as both of our synthetic measures. Since all the dependent variables are standardized, we can directly interpret them in terms of standard deviations. On each of the original measures, responses from the average Shanghai respondent were 14 to 32% of a standard deviation (SD) higher after the announcement of Chen's downfall. The estimated treatment effects from the synthetic measures are even larger: Regardless of which aggregation method we use or our sample choices, the estimates are consistently significant and their magnitudes are as large as about 35% of a standard deviation.

Turning to the implicit measures, our regression estimates suggest that Shanghai residents were significantly less happy (-51.3% SD), less likely to view the court as independent from the government (-41.8% SD), yet at the same time did not use the purported corruption scandal to update their beliefs about how much government officials had gained in the past decade (+0.1 SD but statistically insignificant). The last four models using the synthetic measures also report about 50 to 60% of a standard deviation decline in actual support.

Taken together, the results suggest that the sharp increase in expressed support following the announcement of Chen's dismissal was not founded on a genuine belief in the center's anticorruption initiatives, but was rather accompanied by emotional distress, cynicism toward the legal system and reluctance in updating their perceptions about the state of corruption. To us, the contrasting results from explicit and implicit measures are indicative of the presence of intense preference falsification in Shanghai at that time.

Testing Alternative Explanations

Our main finding suggests that there is a notable divergence in expressed and actual opinion among Shanghai respondents after the purge of Chen Liangyu. Because our design is observational in nature, the validity finding may be subject to a number of alternative explanations. In this section, we conduct several additional tests to evaluate some of the key alternative explanations to our main finding.

First, one might raise the objection that the divergence between explicit and implicit measures that we observe at the aggregate level does not necessarily imply dishonest reporting at the individual level. One possibility, for example, is that a subset of the respondents reacted to the purge *only* by increasing their expressed support and another subset *only* by lowering actual support. If this had been the case, we would still observe opposing movements in attitudes even though both groups were sincere in reporting their preferences. We address this concern in several ways. First, we examine the relationship between residuals from regressions on the explicit and implicit measures. If respondents who have a low opinion about Chen's dismissal do falsify their opinions by reporting a level of expressed support higher than they would otherwise do, we should see a negative correlation between the two sets of residuals (in other words, for those people that the model over-predicts on actual support, it also under-predicts on expressed support), and this negative relationship should be more salient in the treatment period than in the control period. In Figure 4 we fit the relationship between the two sets of residuals (based on Model 6 in Table 2, and we only plot observations in the Shanghai sample) using two separate LOWESS curves for the treatment and control periods. We find a strong negative association between the residuals of the two measures in the treatment period ($\rho_{\text{Treatment}} = -0.178$, $p<0.01$),

but not in the control ($\rho_{\text{Control}} = -0.058$, p<0.57). This pattern suggests that the divergence between expressed and actual opinions did exist at the individual level, and the extent of divergence is much greater during the period in which the propensity to falsify one's preferences is the greatest.

To further verify that the increase in expressed support was, as we have hypothesized, a response to perceived political pressure, we try to look for instruments that exogenously affect the degree of political pressure a respondent faced. One place to look for such an instrument is interviewers' characteristics, which have been found to have tangible impacts on both the quality and the content of responses in other contexts (Blaydes & Gillum, 2013; Davis & Silver, 2003; Schuman & Converse, 1971). For our purpose here, an ideal instrument would be the age of the interviewer. While there is no reason to believe why interviewers' age could directly alter the respondent's genuine political support, it is very plausible that respondents would have greater incentives to demonstrate high expressed support when being interviewed by an older person, who are more likely to be mistaken for a state agent than a younger-looking interviewer.

To formally test this, we again turn to the regression residuals. If the presence of an older interviewer did have some independent effects on expressed support in addition to what have been predicted by all the existing covariates, we would expect a positive correlation between the residuals from expressed support and interviewer's age, and this relationship should likewise be stronger in the treatment period than in the control period. On the right panel of Figure 4, we plot this relationship and report the coefficients from two separate linear regressions for treatment and control periods, respectively. Both visual inspection and the regression estimates confirm that there was a strong, positive correlation between the residuals and interviewers' age

in the Shanghai sample during treatment period ($\beta_{\text{Treatment}} = 0.025$, p<0.05), but the same relationship did not exist during the control period ($\beta_{\text{Control}} = -0.019$, p<0.19).[12] This finding provides further support to the claim that individual-level consideration about political desirability was a key driver behind the marked increase in expressed support following the announcement of Chen's dismissal.

[Figure 4 about here]

A third major concern is that our DID estimates might be biased by influence from other concurrent events. We perform two types of placebo tests to check whether our findings were driven by concurrent interference. First, we create an arbitrary cutoff date for the treatment and control periods and then rerun Model 6 in Table 2. We repeat this exercise for every single day in the survey period.[13] If the observed changes in public attitudes were indeed caused by the purge, then we would expect the coefficients to be maximized or minimized on the *actual* cutoff date or a date close to it.[14] The results from this test are displayed in Figure 5, where we plot the two sets of coefficient estimates and their differences using arbitrary cutoffs from September 12 to October 28. For both expressed and actual support, we see clear U/inverse-U shaped patterns. For expressed support, the coefficient estimate is maximized right on September 26. The turning point for the actual support estimates is somewhat fuzzier, but is still within a week of the

[12] We also look at this relationship for non-Shanghai provinces. For both the treatment and the control periods, the relationship appears to be negative and substantively small.

[13] When displaying the results, we exclude those dates after Oct 28th for presentational considerations. There are indeed very few observations after Oct 28 and none of the results from the placebo tests appear significant.

[14] We should not expect that the coefficient estimates on alternative cutoff dates to be completely insignificant, especially for dates that are close to the actual cutoff. This is because as we incrementally change the cutoff date to an earlier (later) date, we are only making marginal changes to the sample composition by adding to (subtracting from) the treatment sample respondents that were interviewed between the old and the new cutoff dates. Hence we would expect a gradual change in coefficients from this test.

announcement of Chen's purge.[15] The *difference* between the two estimates (highlighted by the red dashed line) is unambiguously maximized on September 26.

[Figure 5 about here]

A related issue is that the changes in Shanghai at the cutoff might be driven by some seasonal factors. While we do not know of any seasonal phenomenon that would create a systematic gap between the expressed and the actual attitudes, it is possible that the **[Happiness]** item in our implicit measures might be influenced by recurring events. To evaluate this possibility, we conduct a second placebo test where we use the CGSS data from the previous year (2005) to estimate the difference in **[Happiness]** before and after the cutoff of September 26 in that year. If there is anything seasonal about the drop in happiness in 2006, we would then expect to see a similar drop in 2005. The results, which are displayed in Table 3, do not support the seasonality hypothesis. No matter which method or sample we use, the estimated coefficient remains small and insignificant. This gives us greater confidence that the dramatic attitudinal change we observe here is very specific to Shanghai as well as the period of the late September of 2006.

[Table 3 about here]

VI. Subgroup Analysis: Who Falsify?

We have so far provided strong evidence that the marked divergence between expressed and actual support caused by the purge of Chen Liangyu is consistent with the phenomenon of

[15] The fuzziness is most likely to be due to sampling variability in our design: The purge was announced on Monday evening and followed by 5 consecutive workdays (September 26 to September 30), which typically had fewer interviews than weekends or national holidays. The relatively small sample size for this period suggests that we might not have a very precise estimate of the public reactions until Oct 1, which is the beginning of a week-long national holiday. The progress during this period was about 5.2 interviews per day. In contrast, during the National holiday (Oct 1 to Oct 7) there was an average of 18.1 interviews per day.

preference falsification. In this section, we relax the assumption that there was uniform response to this event and look at subgroup variations.

Before turning to the results, it is useful to first lay out our conjectures about the basic patterns that might emerge. Generally speaking, two conditions are needed for falsification behavior to arise. First and foremost, preference falsification requires a deviation in underlying beliefs from the dominant discourse (otherwise there would be nothing to falsify on). In an authoritarian context where falsification is related to political considerations, a key determinant for developing unorthodox political beliefs is the access to *alternative sources of information* about politics that helps the recipients resist and counter-argue the dominant political discourse (Geddes & Zaller, 1989; Zaller, 1992). Second, having developed a deviant thought, the individual must also have an *incentive* to conceal it. Past research on social desirability suggest that the incentives to do so can come from (1) an intrinsic need to maintain certain desirable self-image and/or (2) extrinsic costs and benefits for certain types of responses (Paulhus, 1984; Phillips & Clancy, 1972; Tourangeau, Rips, & Rasinski, 2000). In our case, the extrinsic costs and benefits are obviously the more salient consideration. Since punishment on undesirable political expressions will generally come from the state, we expect that those who had more to lose from state sanctions would to be more likely to express supportive attitudes, conditional on the same underlying beliefs.

With these conjectures in mind, we turn to the results at the subgroup level. For each subgroup, we run two regressions for expressed and actual support using the same specifications as Model 6 in Table 3, and map the key causal estimates $\hat{\delta}_{\text{Expressed Support}}$ and $\hat{\delta}_{\text{Actual Support}}$ onto Figure 6. Each circle in the diagram corresponds to a subgroup, with the size of the circle

proportional to the size of that group in the Shanghai sample. We shade the upper half of the circle in grey if $\hat{\delta}_{\text{Expressed Support}}$ is statistically significant at the 90% level or above. Likewise, the bottom half of the circle is shaded in grey when $\hat{\delta}_{\text{Actual Support}}$ is statistically significant.

[Figure 6 about here]

We first draw attention to the upper left quadrant, what we might call the "falsifiers" quadrant, as the divergence between expressed and actual support is greater for these groups than the sample average (in other words, more intense falsification). Four groups are located in this quadrant: (1) wealthy individuals (the top 20% of the local income percentile),[16] (2) frequent internet users,[17] (3) those with college-level education or above,[18] and (4) state sector employees. Memberships in the first three groups are highly correlated.[19] Together, they represent a segment of the Chinese society that is well informed, intelligent, and with relatively high socioeconomic status.

Based on our conjectures on information access, it is not entirely surprising that these groups distrust the official anti-corruption narrative the most. Individuals with higher levels of education are more capable of critically processing official information. The economically better-off class probably had extra sympathy for Chen (note the very negative coefficient on actual support for this group) as many of them might have made their fortune during the tenure of Chen, who was

[16] We construct the local income percentile by ranking individuals within a province by their reported total annual income (sum of wage, rent, interest, investment…etc). Those who did not report their income were not ranked. We created a separate subgroup for those who did not report their income.

[17] Frequent internet users are defined as those who use internet on a *daily* basis.

[18] We define four education categories: "Primary or below", "junior high", "senior high" and "college or above". The "college or above" category includes people who have completed tertiary education with a diploma from junior college (*dazhuan*) or a bachelor degree from a university (*benke*). 12.10% of the respondents in the national sample and 23.25% in the Shanghai sample fall into this category.

[19] The pair-wise polychoric correlation is 0.49 between top income and internet usage, 0.46 between top income and college education, and 0.75 between college education and internet usage.

known for his ability in economic management. Frequent internet users have better access to overseas analyses that saw Chen's dismissal as the result of a power struggle. However, it is somewhat counterintuitive to observe that these groups are also among those that tried hardest to hide their actual opinion and the highest level of support for the Center after the dismissal. In particular, the behaviors of the internet users were in sharp contrast to their usual critical and activist style as documented by many existing studies (King et al., 2013; Yang, 2013). Here, our second conjecture can provide a possible explanation for these puzzling behaviors: These groups, by virtue of their better economic conditions, typically had more to lose from state sanctions than those from the lower socioeconomic strata. At the same time, their intelligence and information also enabled them to gain a better understanding of the risks associated with expressing dissident views at a politically sensitive moment. As a result, while members of these groups might be outspoken at times when speaking up entailed little political cost, they can be extremely cautious about the views they express when the political atmosphere is tightened and the stakes are high.

State-sector employees make up the fourth group in the "falsifiers' quadrant". As system insiders, people in this group are probably too familiar with the vicissitudes of Chinese politics to fully believe in any official narrative. However, their dependence on the state for careers and privileges also provide them with the strongest incentives to hide their real thoughts and demonstrate loyalty at a highly sensitive political moment.

By comparison with state sector employees, private sector employees are less dependent on the state and thus have fewer incentives to feign a supportive attitude. The results confirm this conjecture: The estimate for private sector employees is located toward the lower-left corner. Like state-sector employees, this group also displays a significant decline (of even larger

magnitude) in actual support, yet, unlike the former, there was no corresponding increase in expressed support.

Another group of people often considered to be political insiders are CCP members. Their estimates, however, are statistically insignificant along both dimensions (plotted toward the right of the graph). This may partly be due to the relatively small number of them in the sample but suggest there is a greater level of heterogeneity among CCP members than state sector employees.

VII. Inter-temporal Dynamics

We have shown the widespread presence of preference falsification among Shanghai respondents and the variations in the degree of falsification across subgroups. The question that follows is: What would happen to the preference falsification behavior over an extended time period?

To help understand the dynamics of preference falsification over time, we start with what are logically three scenarios our case of preference falsification might unfold. First, it could persist for an extended period. In this scenario, we would continue to see an elevated level of expressed support despite deterioration in underlying beliefs. Alternatively, as Chen's dismissal fade in salience and memory, the national leadership might over time be able to win more of the public over to the anti-corruption narrative that Chen was indeed guilty of corruption—something that it could not achieve in the short-run. In consequence the level of actual support would recover. A third possibility would be for the respondents to remain skeptical about the real motives behind

Chen's dismissal and yet the pressure for falsification receded. As a result, respondents would become more willing to share their true attitudes than previously and we would expect the level of expressed support to decline over time.

To investigate which of these scenarios is closer to reality, we analyze data from two additional surveys that were conducted about one year apart from each other. The Chinese Social Survey (CSS06), by the Institute of Sociology, Chinese Academy of Social Sciences, was completed between April and June 2006 (about three months before Chen's dismissal), and the World Value Survey (WVS07) was administered between March and May in 2007. We focus on two sets of questions on general political trust in these two surveys.

[Trust in Central Government]:

- "(CSS06) *How much do you trust the central government? (1=Very little, 2=Not much, 3=Somewhat, 4=A great deal)*"
- "(WVS07) *How much do you trust your government in Beijing? (1=Not at all, 2=Not much, 3=Somewhat, 4=A great deal)*"

[Trust in the Party:]

- "(CSS06) *Do you agree with the following statement: Our Party and government are capable of properly handling problems that our country is currently facing" (1=Strongly Disagree, 2=Disagree, 3=Agree, 4=Strongly agree)*"
- "(WVS07) *How much do you trust political parties? (1=Not at all, 2=Not much, 3=Somewhat, 4=A great deal)*"

The wordings for the first set of questions on trust in central government are almost identical, but there are some notable differences in the questions on Party trust. The WVS07 asked about trust in political parties in general, rather than a specific party. Yet, given the CCP's omnipresence in China, it is reasonable to expect that interviewees will have it in mind when responding to this question. To make the responses more comparable across different surveys, we use the same DID approach and implement entropy balancing to reweight observations on common observables (*Age, Gender, Education, Sector of employment, Employment status, Managerial position*). We also include controls for region-treatment interactions to take into account systematic between-survey differences across regions.[20]

Table 4 presents the results on attitudinal change. Our key coefficient of interest here is again the interaction term between treatment and Shanghai dummies, which tells us how much the attitudes of Shanghai residents changed between the two surveys relative to the rest of China. In all models and for both dependent variables, we see a significant drop in political trust that amounts to about 5 to 7 percent of the overall scale. The decline is statistically significant as well as large, especially when compared to other social and economic determinants of political trust in existing research. This notable decline supports the third scenario: Negative feelings Shanghai residents held toward the Center were not mitigated by the passage of time but became more evident as political pressure receded.

[Table 4 about here]

VIII. Concluding Remarks

[20] The three regional dummies we have are for North, East and ethnic autonomous regions. We interact them with the treatment dummy to take into account all region-specific heterogeneities between surveys. The DID estimates are therefore based on within-region comparisons.

Much of the knowledge creation in social science relies on self-reported data, which are prone to biases and human manipulations. In this study our innovative quasi-experimental design has allowed us to systematically document the presence of preference falsification under authoritarian systems and to explore its determinants. We show that, contrary to conventional wisdom, preference falsification in a relatively open authoritarian system like China's is a contingent phenomenon that rises and falls in response to changes in the political atmosphere. We also find that the inter-group variations in the incidence of preference falsification are best captured by the combination of information exposure and vulnerability to state sanctions.

Although this study focuses on a unique event in Chinese politics, we believe our findings are generalizable to similar cases in authoritarian settings and offer important insight into the dynamics of authoritarianism. Political purges may not be daily occurrences but they are by no means alien to those living in former Soviet Union, China (recent cases include those of Bo Xilai, Zhou Yongkang and Ling Jihua), North Korea (Jang Song Thaek) and many other authoritarian systems. Since Chen's dismissal occurred in times of relatively halcyon political times and still produced so much impact on public attitudes, we can imagine how much more subjects in these countries would be watching their words in times of Great Purges and massive political campaigns such as the Anti-Rightist Campaign, the Great Leap Forward, and the Cultural Revolution.

More generally, our study sheds light on the interplay between the two key challenges facing authoritarian leaders: (1) securing consent among the ruling elites and (2) preventing a popular uprising from the masses (Boix & Svolik, 2007; Bueno de Mesquita, Smith, Siverson, & Morrow, 2005; Svolik, 2012). Our finding that the broad public reacted strongly to the dismissal of Chen Liangyu, a non-elected local politician, shows that the public under authoritarian regimes possess

substantial knowledge about high politics and many also felt strongly about such political developments.[21] By underscoring the importance of mass opinion concerning elite politics in China, our study thus links up the twin challenges facing authoritarian leaders and elucidates why authoritarian leaders in China and elsewhere care so much about public opinion even on matters of elite politics.

Last but not the least, our findings remind us of the need to use survey data from authoritarian settings with a heavy dose of caution and not to take the responses at their face value. Significant public events may alter the relative cost and benefit of expressing certain attitudes (Berinsky, 2002) and affect the meaning of questions by changing the relative salience of various considerations respondents have in their minds (Zaller & Feldman, 1992). As our analysis shows, a naïve focus on the expressed measures of support alone would have led researchers to draw erroneous conclusions about Shanghai's public sentiment toward the central authorities in the aftermath of Chen's dismissal. Only by incorporating information from the related but non-sensitive survey items were we able to discover the widespread resentment beneath the surface of unanimous support. Our study suggests that great care is needed in interpreting results from those surveys and deep contextual knowledge about the timing and the nature of the survey can help substantially in coping with biases and uncovering hidden patterns.

21 International relations scholars are among the first to recognize this linkage. See the literature on audience cost (Fearon, 1994; Trachtenberg, 2012; Weeks, 2008; Weiss, 2013).

References

Arendt, Hannah. (1972). *Crises of the Republic: Lying in Politics; Civil Disobedience; On Violence; Thoughts on Politics and Revolution*: Mariner Books.

Berinsky, Adam J. (2002). Political Context and the Survey Response: The Dynamics of Racial Policy Opinion. *The Journal of Politics, 64*(02), 567-584. doi: doi:10.1111/1468-2508.00140

Bernstein, Robert, Chadha, Anita, & Montjoy, Robert. (2001). Overreporting Voting: Why It Happens and Why It Matters. *Public Opinion Quarterly, 65*(1), 22-44. doi: 10.1086/320036

Blaydes, Lisa. (2010). *Elections and Distributive Politics in Mubarak's Egypt*: Cambridge University Press.

Blaydes, Lisa, & Gillum, Rachel M. (2013). Religiosity-of-Interviewer Effects: Assessing the Impact of Veiled Enumerators on Survey Response in Egypt. *Politics and Religion*, 1-24.

Boix, Carles, & Svolik, Milan. (2007). The Foundations of Limited Authoritarian Government: Institutions and Power-Sharing in Dictatorships. *Journal of Politics, 75*(2), 300-316.

Bradburn, Norman.M., Sudman, Seymour., & Blair, Edward. (1979). *Improving interview method and questionnaire design*: Jossey-Bass.

Bueno de Mesquita, Bruce., Smith, Alastair, Siverson, Randolph M., & Morrow, James D. (2005). *The Logic of Political Survival*: MIT Press.

Chen, Jie. (2004). *Popular Political Support in Urban China*: Stanford, California: Stanford University Press.

Corstange, Daniel. (2010). *Vote Buying under Competition and Monopsony: Evidence from a List Experiment in Lebanon*. Paper presented at the Annual Conference of American Political Science Association, Washington, D.C.

Darden, Keith, & Grzymala-Busse, Anna. (2006). The Great Divide. *World Politics, 59*(1), 83-115.

Davis, Darren W., & Silver, Brian D. (2003). Stereotype Threat and Race of Interviewer Effects in a Survey on Political Knowledge. *American Journal of Political Science, 47*(1), 33-45. doi: 10.1111/1540-5907.00003

Fazio, Russell H., & Olson, Michael A. (2003). Implicit Measures in Social Cognition Research: Their Meaning and Use. *Annual Review of Psychology, 54*, 297-327.

Fearon, James D. (1994). Domestic Political Audiences and the Escalation of International Disputes. *The American Political Science Review, 88*(3), pp. 577-592.

Fong, Leslie. (2004). Leadership Dispute over China Growth, *Strait Times*. Retrieved from http://straitstimes.asia1.com.sg.

Geddes, Barbara, & Zaller, John. (1989). Sources of Popular Support for Authoritarian Regimes. *American Journal of Political Science, 33*(2), 319-347.

Gilens, Martin, Sniderman, Paul M., & Kuklinski, James H. (1998). Affirmative Action and the Politics of Realignment. *British Journal of Political Science, 28*(1), 159-183. doi: 10.2307/194161

Gonzalez-Ocantos, Ezequiel, de Jonge, Chad Kiewiet, Meléndez, Carlos, Osorio, Javier, & Nickerson, David W. (2012). Vote Buying and Social Desirability Bias: Experimental Evidence from Nicaragua. *American Journal of Political Science, 56*(1), 202-217. doi: 10.1111/j.1540-5907.2011.00540.x

Hainmueller, Jens. (2011). Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis*. doi: 10.1093/pan/mpr025

Havel, Vaclav. (1990). *The Power of the Powerless: Citizens Against the State in Central-eastern Europe*: M.E. Sharpe.

Holbrook, Allyson L., & Krosnick, Jon A. (2010). Social desirability bias in voter turnout reports: Tests using the item count technique. *Public Opinion Quarterly, 74*(1), 37-67. doi: 10.1093/poq/nfp065

Kane, James G., Craig, Stephen C., & Wald, Kenneth D. (2004). Religion and Presidential Politics in Florida: A List Experiment*. *Social Science Quarterly, 85*(2), 281-293. doi: 10.1111/j.0038-4941.2004.08502004.x

Karp, Jeffrey A., & Brockington, David. (2005). Social Desirability and Response Validity: A Comparative Analysis of Overreporting Voter Turnout in Five Countries. *Journal of Politics, 67*(3), 825-840. doi: 10.1111/j.1468-2508.2005.00341.x

Kern, Holger Lutz, & Hainmueller, Jens. (2009). Opium for the Masses: How Foreign Media Can Stabilize Authoritarian Regimes. *Political Analysis*. doi: 10.1093/pan/mpp017

King, Gary, Pan, Jennifer, & Roberts, Margaret E. (2013). How Censorship in China Allows Government Criticism but Silences Collective Expression. *American Political Science Review, 107*, 326-343.

Kuklinski, James H., Cobb, Michael D., & Gilens, Martin. (1997). Racial Attitudes and the "New South". *The Journal of Politics, 59*(02), 323-349. doi: doi:10.2307/2998167

Kuklinski, James H., Sniderman, Paul M., Knight, Kathleen, Piazza, Thomas, Tetlock, Philip E., Gordon, R. Lawrence, & Mellers, Barbara. (1997). Racial Prejudice and Attitudes Toward Affirmative Action. *American Journal of Political Science, 41*(2), 402-419. doi: 10.2307/2111770

Kuran, Timur. (1991). Now Out of Never: The Element of Surprise in the East European Revolution of 1989. *World Politics, 44*(1), 7-48. doi: 10.2307/2010422

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

Kuran, Timur. (2011). The Politics of Revolutionary Surprise. Retrieved July 19 2014, from http://www.project-syndicate.org/commentary/the-politics-of-revolutionary-surprise

Kurzman, Charles. (2009). *The Unthinkable Revolution in Iran*: Harvard University Press.

Lewis-Beck, Michael S., Tang, Wenfang, & Martini, Nicholas F. (2013). A Chinese Popularity Function: Sources of Government Support. *Political Research Quarterly*.

Li, Cheng. (2007). Was the Shanghai Gang Shanghaied? The Fall of Chen Liangyu and the Survival of Jiang Zemin's Faction. *China Leadership Monitor*(20).

Linz, Juan .J. (2000). *Totalitarian and Authoritarian Regimes*: Lynne Rienner Publishers.

Lohmann, Susanne. (1994). The Dynamics of Informational Cascades: The Monday Demonstrations in Leipzig, East Germany, 1989–91. *World Politics, 47*(01), 42-101. doi: doi:10.2307/2950679

Lü, Xiaobo. (2014). Social Policy and Regime Legitimacy: The Effects of Education Reform in China. *American Political Science Review, 108*(2).

Malesky, Edmund, Schuler, Paul, & Tran, Anh. (2012). The Adverse Effects of Sunshine: A Field Experiment on Legislative Transparency in an Authoritarian Assembly. *American Political Science Review, 106*(04), 762-786. doi: doi:10.1017/S0003055412000408

Manion, Melanie. (2004). *Corruption by Design*: Harvard University Press.

Manion, Melanie. (2010). A Survey of Survey Research on Chinese Politics. In Allen Carlson, Mary E. Gallagher, Kenneth Lieberthal & Melanie Manion (Eds.), *Contemporary Chinese Politics: New Sources, Methods and Field Strategies* (pp. 181-199): Cambridge University Press.

Meng, Tianguang, Pan, Jennifer, & Yang, Ping. (2014). Conditional Receptivity to Citizen Participation: Evidence From a Survey Experiment in China. *Comparative Political Studies*. doi: 10.1177/0010414014556212

Patel, David Siddhartha. (2012). *Roundabouts and Revolutions: Public Squares, Coordination and the Diffusion of the Arab Uprisings*. Department of Government, Cornell University.

Paulhus, Delroy L. (1984). Two-Component Models of Socially Desirable Responding. *Journal of Personality and Social Psychology, 46*(3), 598-609.

Phillips, Derek L., & Clancy, Kevin J. (1972). Some Effects of "Social Desirability" in Survey Studies. *American Journal of Sociology, 77*(5), 921-940. doi: 10.2307/2776929

Powell, Richard J. (2013). Social Desirability Bias in Polling on Same-Sex Marriage Ballot Measures. *American Politics Research, 41*(6), 1052-1070. doi: 10.1177/1532673x13484791

Rose, Richard. (2007). Perspectives on Political Behavior in Time and Space. In Russell .J. Dalton & Hans-Dieter Klingemann (Eds.), *Oxford Handbook of Political Behavior*: Oxford University Press.

Rose, Richard, Mishler, William, & Munro, Neil. (2011). *Popular Support for an Undemocratic Regime*. New York: Cambridge University Press.

Schuman, Howard, & Converse, Jean M. (1971). The Effects Of Black And White Interviewers On Black Responses In 1968. *Public Opinion Quarterly, 35*(1), 44-68. doi: 10.1086/267866

Silver, Brian D., Anderson, Barbara A., & Abramson, Paul R. (1986). Who Overreports Voting? *The American Political Science Review, 80*(2), 613-624. doi: 10.2307/1958277

Streb, Matthew J., Burrell, Barbara, Frederick, Brian, & Genovese, Michael A. (2008). Social Desirability Effects and Support for a Female American President. *Public Opinion Quarterly, 72*(1), 76-89. doi: 10.1093/poq/nfm035

Sulek, Antoni. (1994). Systemic Transformation and the Reliability of Survey Research: Evidence from Poland. *Polish Sociological Review*(106), 85-100. doi: 10.2307/41274531

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

Tang, Wenfang. (2005). *Public Opinion and Political Change in China*: Stanford, California: Stanford University Press.

Tessler, Mark. (2002). Islam and Democracy in the Middle East: The Impact of Religious Orientations on Attitudes toward Democracy in Four Arab Countries. *Comparative Politics, 34*(3), 337-354. doi: 10.2307/4146957

Tourangeau, Roger., Rips, Lance .J., & Rasinski, Kenneth. (2000). *The Psychology of Survey Response*: Cambridge University Press.

Trachtenberg, Marc. (2012). Audience Costs: An Historical Analysis. *Security Studies, 21*(1), 3-42.

Treisman, Daniel. (2011). Presidential Popularity in a Hybrid Regime: Russia under Yeltsin and Putin. *American Journal of Political Science, 55*(3), 590-609. doi: 10.1111/j.1540-5907.2010.00500.x

Webb, Eugene J., Campbell, Donald T., Schwartz, Richard, & Scechrest, Lee. (1966). *Unobtrusive Measures: Nonreactive Research in the Social Sciences*. Chicago: Rand McNally.

Wedeen, Lisa. (1999). *Ambiguities of Domination: Politics, Rhetoric, and Symbols in Contemporary Syria*: University of Chicago Press.

Weeks, Jessica L. (2008). Autocratic Audience Costs: Regime Type and Signaling Resolve. *International Organization, 62*(1), pp. 35-64.

Weiss, Jessica Chen. (2013). Authoritarian Signaling, Mass Audiences, and Nationalist Protest in China. *International Organization, 67*(1), 1-35.

Welsh, William A. (Ed.). (1981). *Survey Research and Public Attitudes in Eastern Europe and the Soveit Union*: New York: Pergamon Press.

Yang, Guobin. (2013). *The Power of the Internet in China: Citizen Activism Online*: Columbia University Press.

Zaller, John. (1992). *The Nature and Origins of Mass Opinion*: Cambridge University Press.

Zaller, John, & Feldman, Stanley. (1992). A Simple Theory of the Survey Response: Answering Questions versus Revealing Preferences. *American Journal of Political Science, 36*(3), pp. 579-616.

Zhao, Dingxin. (2009). Mandate of Heaven and Performance Legitimacy in Historical and Contemporary China. *American Behavioral Scientist, 53*, 416-433e.

Table 1 Predicted Directions of Attitudinal Change for Respondents with Different Attitudes toward the Purge

	Happiness	Judicial Independence	Official as the Greatest Beneficiary
Supporters (Believed in Anti-corruption)	+	No Change	+
Opponents (Sympathized with Chen)	-	-	No Change

Table 2 Overall Effects of the Purge from Explicit and Implicit Measures

Panel 1: Explicit Measures	Compliance	Government over Law	Development over Democracy	Synthetic I: Simple Average		Synthetic II: 1st Principal Component	
	(1) Full sample	(2) Full sample	(3) Full sample	(4) Full sample	(5) Before Nov 1 only	(6) Full sample	(7) Before Nov 1 only
Treatment x Shanghai	0.294* (0.169)	0.143 (0.184)	0.321* (0.185)	**0.360*** **(0.185)**	**0.355*** **(0.186)**	**0.351*** **(0.185)**	**0.345*** **(0.186)**
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5046	5046	5046	5046	5000	5046	5000

Panel 2: Implicit Measures	Happiness	Judicial Independence	Beneficiary Group	Synthetic I: Simple Average		Synthetic II: 1st Principal Component	
	(1) Full sample	(2) Full sample	(3) Full sample	(4) Full sample	(5) Before Nov 1 only	(6) Full sample	(7) Before Nov 1 only
Treatment x Shanghai	-0.513*** (0.168)	-0.418** (0.197)	0.105 (0.216)	**-0.497**** **(0.249)**	**-0.490**** **(0.249)**	**-0.591**** **(0.235)**	**-0.584**** **(0.235)**
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5046	5046	5046	5046	5000	5046	5000

Note: Analyses are based on 5 multiply imputed datasets. The demographic controls include *age, gender, sector of employment, Party membership, college degree, ethnicity, marital status, household registration status (hukou), employment status* and *military experience*. Robust standard errors clustered at survey location levels are reported in parentheses. Entropy balancing is implemented in all models.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3 Placebo Regressions on Happiness Using Data from CGSS 2005

	(1) Shanghai only	(2) Shanghai only	(3) Shanghai only	(4) Full Sample (DID)	(5) Full Sample (DID)	(6) Full Sample (DID)
Treatment	**-0.058**	**-0.059**	**0.001**	-0.066	-0.034	0.120
	(0.204)	(0.149)	(0.139)	(0.131)	(0.120)	(0.269)
Treatment x Shanghai				**0.068**	**0.065**	**-0.090**
				(0.207)	(0.195)	(0.190)
District FE	No	Yes	Yes	Yes	Yes	Yes
Date FE	-	-	-	No	Yes	Yes
Demographics	No	No	Yes	No	No	Yes
Observations	400	400	400	5623	5623	5623

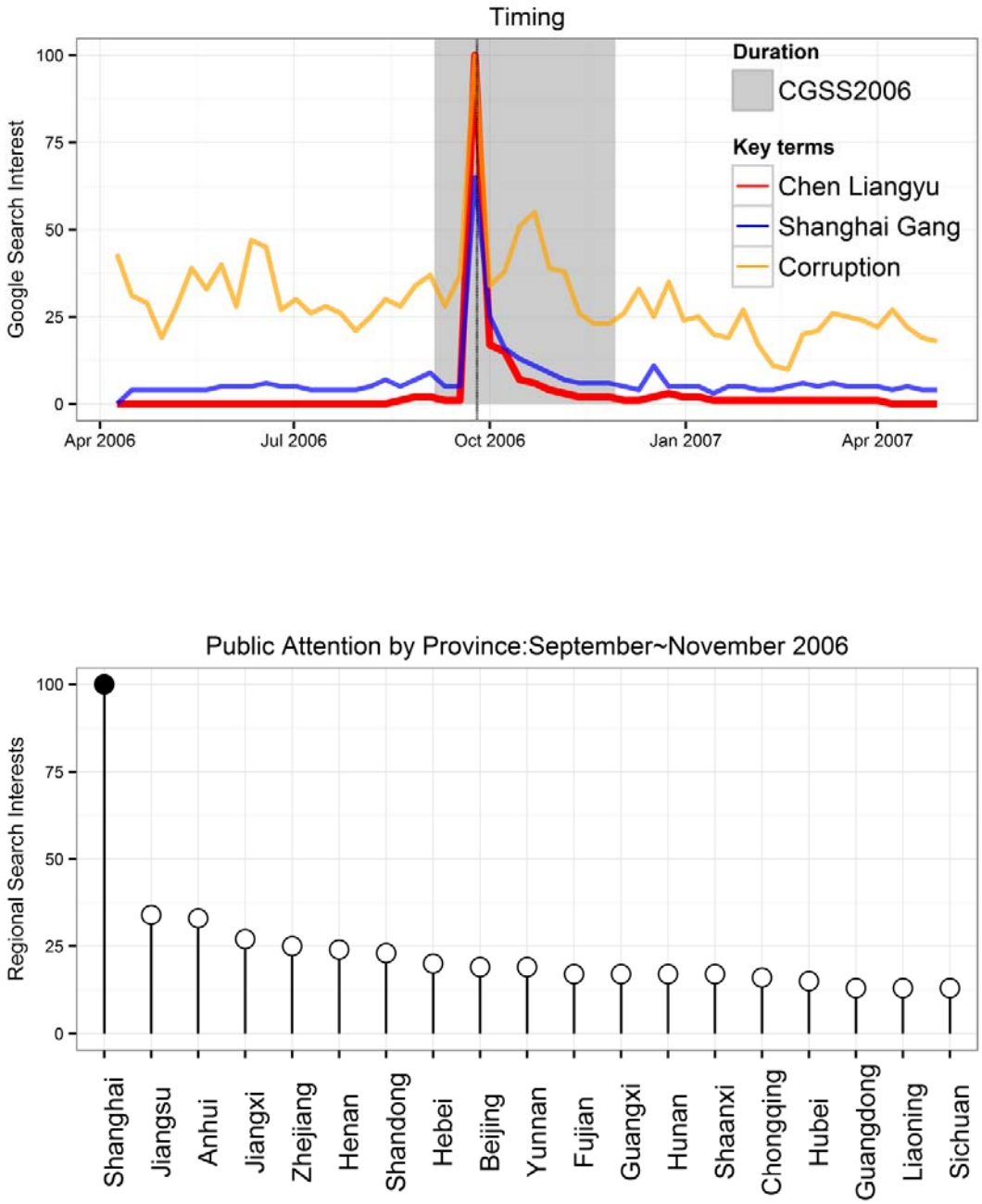
Note: The dependent variable is standardized response from the [**Happiness]** question. The demographic controls include *age, gender, sector of employment, Party membership, college degree, ethnicity, marital status, household registration status (hukou), employment status* and *military experience*. Robust standard errors clustered at survey location levels are reported in parentheses. Entropy balancing is implemented in all models. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4 Decline in Trust as the Pressure on Falsification Diminished

	Trust in the central government				Trust in the Party			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	-0.284***	-0.277***	-0.229**	-0.233**	-0.165*	-0.167*	-0.186*	-0.211**
x Shanghai	(0.094)	(0.093)	(0.097)	(0.104)	(0.098)	(0.098)	(0.101)	(0.107)
Regional FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Regional trends	No	No	Yes	Yes	No	No	Yes	Yes
Observations	9076	9076	9076	9076	9076	9076	9076	9076

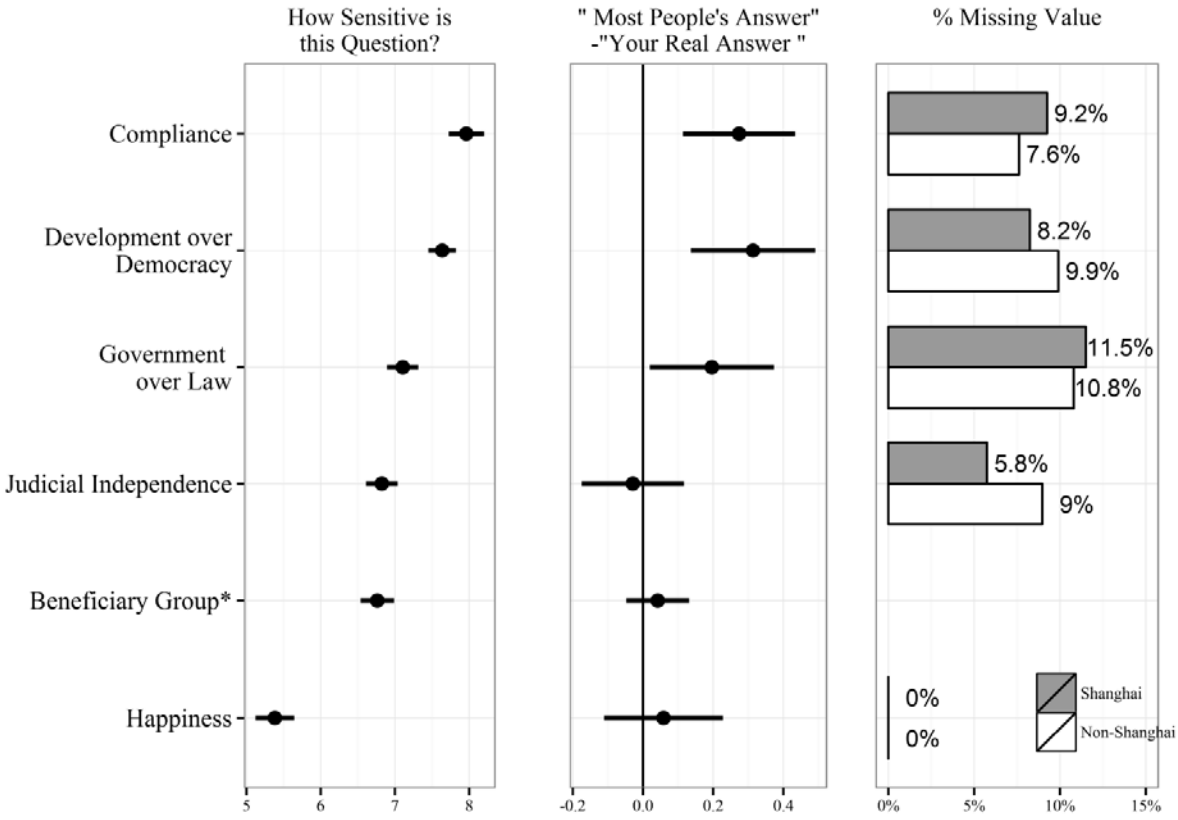
Notes: Analyses are based on 5 multiply imputed datasets. The demographic covariates include *age*, *female*, *sector of employment*, *college degree* and *profession*. We include three regional dummies for North, East and ethnic autonomous regions. Robust standard errors are reported in parentheses. As clustering at provincial level yields smaller standard errors (due to negative intra-cluster correlation), we report the standard errors without clustering. Entropy balancing is implemented in all models.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure 1: Temporal and Spatial Distributions of Public Attention to the Purge



Note: The three lines on the first graph represent the Google search interest from September to November in 2006. The values for the key words "Corruption" and "Shanghai Gang" have been scaled so that they are comparable. The shaded area indicates the duration of the survey.

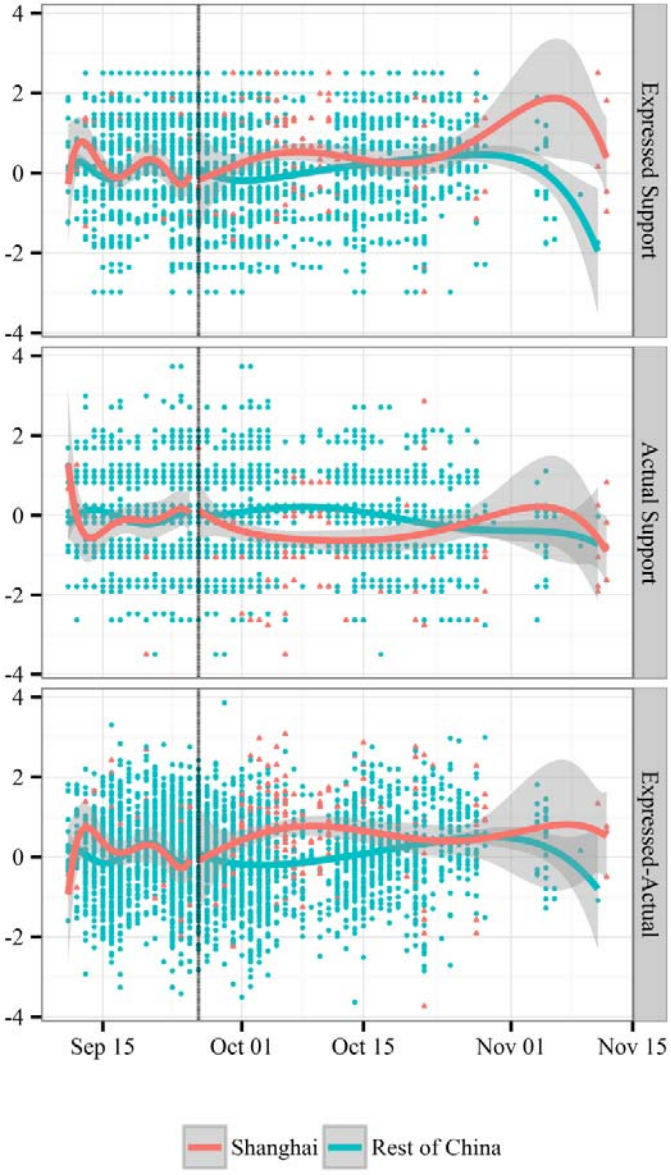
Figure 2: Evaluating the Relative Sensitivity of the Explicit and Implicit Measures



Notes: For each measure (indicated on the y axis), the left panel displays the average level of perceived question sensitivity reported by the experiment subjects (n=311), with a large number indicating high level of perceived sensitivity. The right panel reports the standardized difference (calculated by dividing the mean difference by standard deviation) in answers between the treatment group (n=149), which was instructed to give "the answer most people would give" and the control group (n=162), which was instructed to give "your real answer if there's no external pressure". A large number here indicates possible over-reporting in face of external pressure. In both panels, the circles represent the point estimates and horizontal bars represent the 90% confidence intervals.

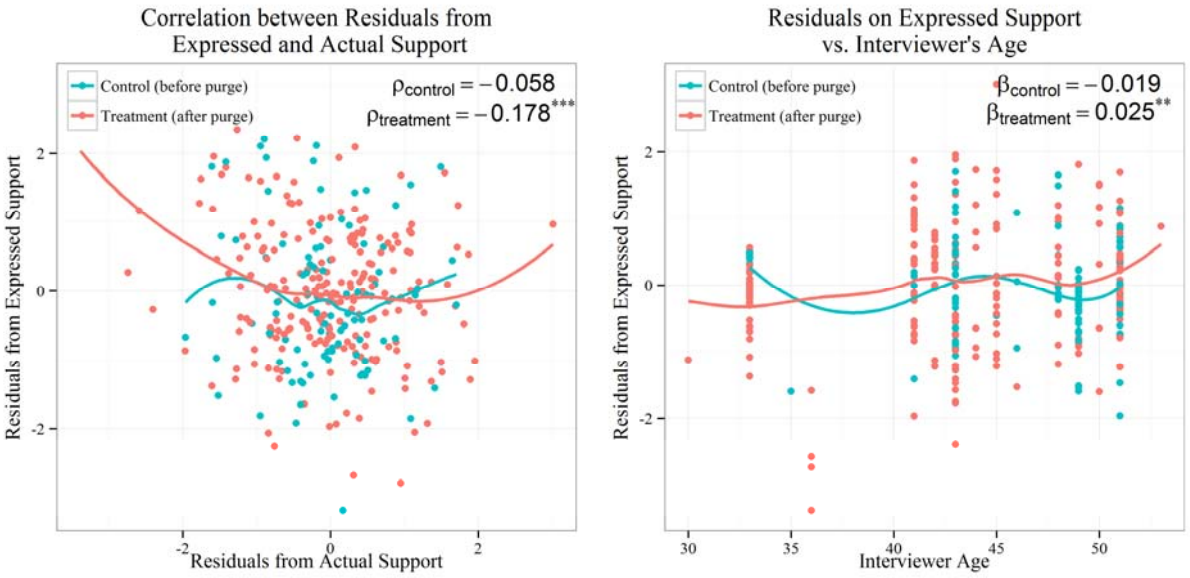
* Since this question does not allow no-response, we do not plot its missing rate on the right panel.

Figure 3 Expressed and Actual Support in Shanghai: Before and After the Purge



Note: Each graph represents the over-time change of the synthetic measures of expressed support (1st principal component), actual belief and falsification in the Shanghai sample. We fit a 5th order polynomial for the sample before and after the purge and the shaded areas represent the 95% confidence interval.

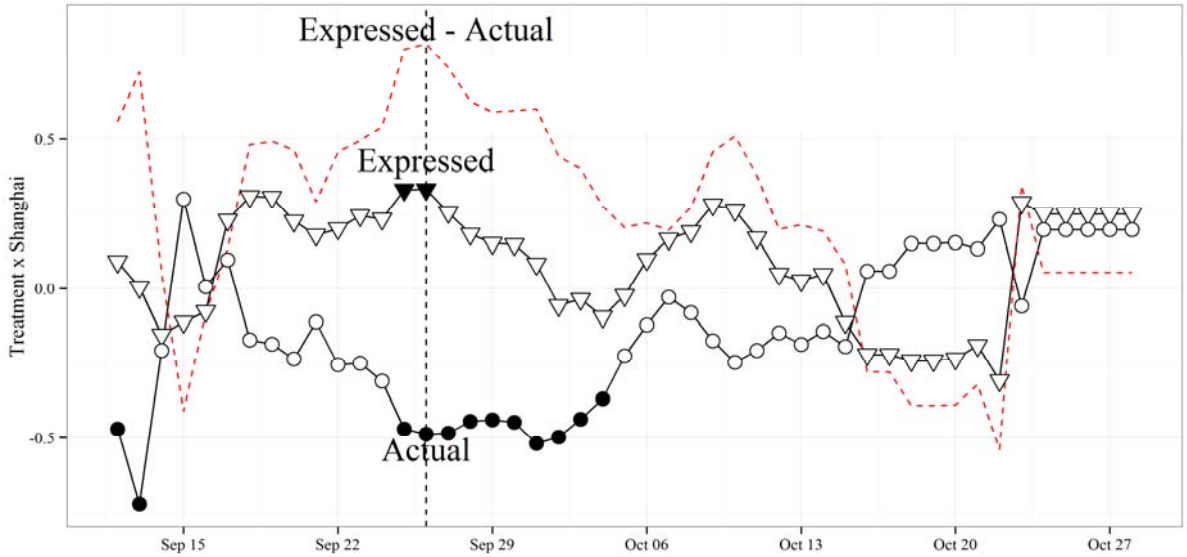
Figure 4: Analysis of Residuals



Note: In each panel, two separate Lowess curves are fitted for the treatment and the control groups in Shanghai. The coefficients on the top right of the first panel on the left is the Pearson correlation for the two residual measures, and those on the second panel are obtained by regression the residuals on interviewers' age.

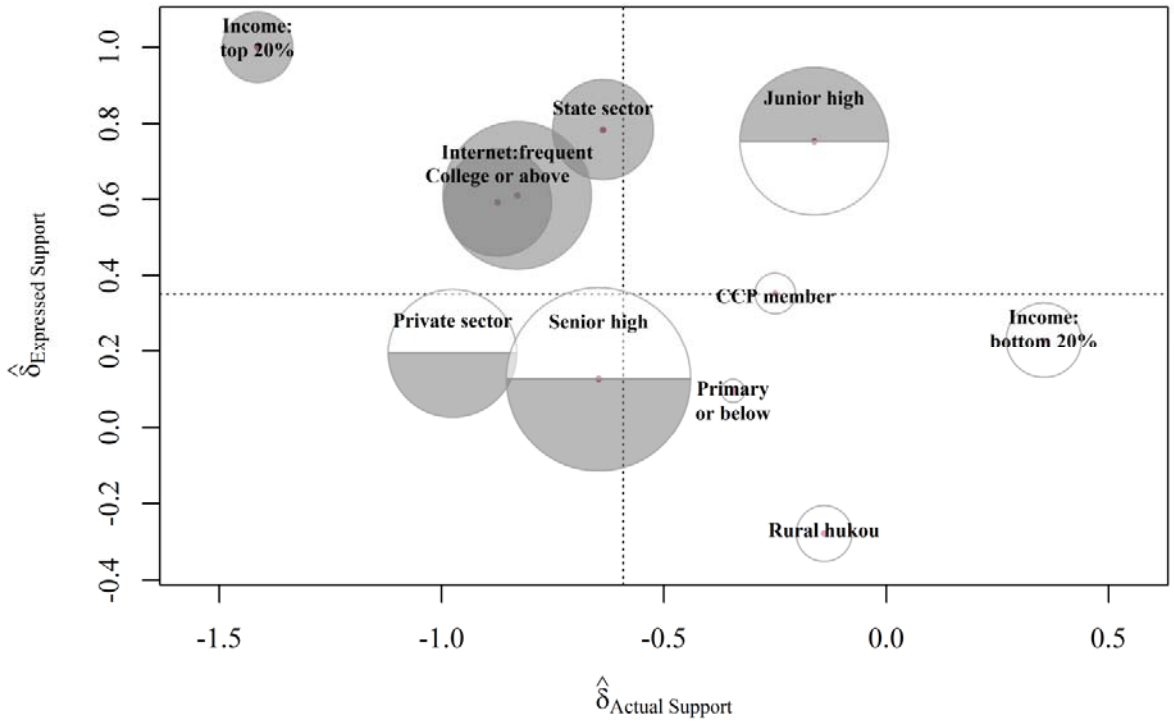
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure 5 Testing Different Cutoff Dates



Note: Each circle/triangle represents the DID estimate for the Shanghai sample using different cutoff dates (indicated on the x axis) for expressed and actual support. Dates with significant coefficient estimates (p<0.1) are colored in black. The red dashed line represents the difference in coefficient estimates between the explicit and actual support. The vertical dashed line marks the actual cutoff on the date of September 26 2006.

Figure 6 Effects of Purge on Expressed and Actual Support: by Subgroups



Notes: Each circle represents a pair of estimated treatment effects (Treatment x Shanghai) on expressed (y axis) and actual support (x axis) for a given subgroup (indicated on the circle). The size of each circle is proportional to the size of the subgroup. Grey shading on the upper half of the circle means indicates a statistically significant (p<0.1) treatment effect in the regression on expressed support. Similarly, shading on the bottom half of the circle indicates a statistically significance treatment effect in the regression on actual support. Full shading indicates significance in both regressions. The dotted vertical and horizontal lines represent the estimates using the *full sample*. All estimates are based on the same specification as Model 6 of Table 2.