

Lying or Believing? Measuring Preference Falsification from a Political Purge in China¹

Junyan Jiang² Dali L. Yang³

Forthcoming, *Comparative Political Studies*

Abstract

Despite its wide usage in explaining political dynamics of nondemocracies, preference falsification remains an empirical myth for students of authoritarian politics due to the challenge of measurement. We offer the first quantitative study of this phenomenon in a nondemocratic setting by exploiting 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 and track their changes before and after the purge. We find that the purge caused 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 preference falsification and conduct a number of robustness checks to rule out alternative explanations. We also show that falsification was most intense among groups that had access to alternative information but were vulnerable to political sanctions.

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 David Samuels, three anonymous reviewers, and the following individuals for helpful comments and suggestions: Michael Albertus, Sofia Alia, Milena Ang, 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. We also thank Shizheng Feng for his assistance with the CGSS data, Zhuang Liu for his help on the online experiment, and Yalin Liu for research assistance.

²PhD candidate, Department of Political Science, University of Chicago

³ William Claude Reavis Professor, Department of Political Science, University of Chicago

I. Introduction

Preference falsification, defined as the act of deliberately misrepresenting one's genuine views and wants under perceived social pressure (Kuran, 1997), exists in a wide range of social and political settings. A large body of research on democratic countries has found that in face-to-face opinion polls, people often try to hide their true attitudes toward certain minority groups (Gilens, Sniderman, & Kuklinski, 1998; Kane, Craig, & Wald, 2004; Kuklinski, Cobb, & Gilens, 1997), overstate support for some candidates, institutions, or political values (Holbrook & Krosnick, 2010; Kuklinski, Sniderman, et al., 1997; Payne, 2010; Powell, 2013; Streb, Burrell, Frederick, & Genovese, 2008), and falsely report their propensity to undertake socially desirable activities (Bernstein, Chadha, & Montjoy, 2001; Cox, Jones, & Juhem, 2014; Holbrook & Krosnick, 2010; Lee & Woodliffe, 2010).

If preference falsification is already common in democracies, it is probably even more pervasive in authoritarian regimes, where legal protection for rights and freedom of speech is either weak or absent. Lacking legitimacy from popular mandates, authoritarian regimes often take a direct interest in what their citizens say and do in politics and routinely sanction those who dare to openly challenge the official discourse (Arendt, 1972; Linz, 2000). To avoid imminent punishment, ordinary citizens under those systems were often compelled to feign support for what they privately oppose (Havel, 1990). Such practice, when adopted by many, distorts public knowledge about the regime's true popularity and undermines coordinated actions among citizens even when a large segment of the society is already privately dissatisfied with the regime. Widespread preference falsification is believed to be an important factor in accounting for the longevity of many authoritarian regimes in Eastern Europe, the former Soviet Union, and the Middle East, as well as for their surprising collapses following seemingly insignificant events

that made information about mass discontent public knowledge (Kuran, 2011; Kurzman, 2009; Lohmann, 1994).

Although many influential theoretical propositions about regime dynamics in dictatorships hinge on the assumption that preference falsification has a constant and pervasive presence in those systems, this assumption is rarely tested against systematic data. As a result, the strongest evidence supporting this assumption so far comes from personal accounts and in-depth ethnographic case studies (Havel, 1990; Wedeen, 1999), which cannot be easily generalized to the population at large. The bulk of the survey-based research on public opinion in authoritarian regimes, however, have mostly avoided a direct discussion of this issue, either by seeking a better research design that achieves 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. Moreover, while powerful experimental techniques are being introduced to the study of authoritarian politics (Malesky, Schuler, & Tran, 2012; Meng, Pan, & Yang, 2014), the political authorities in authoritarian regimes typically place heavy restrictions on the scope and the types of questions allowed, and politically sensitive topics are almost always off the table.⁴

In this paper, we adopt a new, natural-experiment based approach to measure the extent of preference falsification from regular opinion surveys in settings where conducting controlled experiments are difficult or infeasible. The general approach requires two key elements: (1) unexpected occurrence of a significant public event that exogenously alters the relative costs and benefits associated with expressing certain politically sensitive attitudes, and (2) a survey that

⁴ In the case of China, for example, it is required by law that surveys with political content must be placed on record (beian) with the supervising authority and carried out only by institutions with the necessary qualifications. Certain topics are prohibited and violations can lead to the revoking of survey license.

intersects with the occurrence of the event and contains differentially sensitive questions about the attitude of interest. With the assumption that preference falsification is less severe in responses to the less sensitive questions than the more sensitive ones, a comparison of the direction and degree of changes in survey responses to the two sets of questions will allow us to estimate the extent of preference falsification in the aftermath of the event.

We use this approach to study public reactions 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 led by a major Chinese university, was being conducted across China and Shanghai was included in the sampling frame. The occurrence of the purge divided the sample into two groups, a “treatment” group that was interviewed after the purge and a “control” group that was interviewed before that. Although in no way designed for our study, the survey included both questions that explicitly ask respondents’ *political support* for the state, and those 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 political support* from a number of survey items, and track their changes before and after the purge. With the assumption that the progression of the survey is largely uncorrelated with respondents’ characteristics—a point we will demonstrate later—this quasi-experimental setting allows us to assess how respondents in Shanghai reacted to this major political event in terms of both their expressed attitudes and actual beliefs.

We find that in the aftermath of the purge the average Shanghai respondents became much more likely to give affirmative answers to sensitive survey questions explicitly linked to regime

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, all suggesting that they had little genuine faith 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 a number of tests to rule out alternative explanations.

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 potential political sanctions. The variations we found are largely consistent with the conjecture: preference falsification was most intense among the wealthy, the highly educated, frequent internet users, public-sector employees, and the generation who spent their formative years during the Cultural Revolution.

Our study contributes to several strands of research. First and foremost, it contributes to the literature on comparative mass political behavior by providing the first set of estimates on the extent of preference falsification in an authoritarian setting. Contrary to the notion that preference falsification is no longer a serious issue for studies of authoritarian states in the age of information because of the scale and anonymity associated with online communication (Farrell, 2012), we show that it is still evident and widespread in off-line expression even in a relatively open system like China's. When it comes to important political issues on which the Party state has a clear position, ordinary citizens still do not dare to publicly reveal their disagreement with the dominant discourse. Meanwhile, our study also updates the classical perspectives on authoritarian politics by showing that preference falsification is better construed as a *variable*

than a constant. The propensity of the citizens to misrepresent their true opinions is not only a function of the overall political environment but also has important individual-level determinants.

Second, our study sheds light on the debate about the utility of employing social surveys to measure public opinion in non-democracies. Students of comparative politics have so far remained 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). The finding that respondents do deliberately overstating their regime support in politically sensitive moments confirms the former group's concern about the reliability of authoritarian surveys. On the other hand, however, we argue that it is possible to overcome the potential biases and extract meaningful information about public sentiment from those surveys. Our finding of declining actual support among Shanghai respondents in the aftermath of the purge illustrates how hidden political dissent can be identified from conventional surveys by incorporating substantive knowledge about the survey context with a carefully chosen research design.

By assessing the impact of a political purge on overt and actual political support, our study also contributes to the 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). While existing studies emphasize the role of economic performance in sustaining regime support, 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

politically motivated could produce intense emotional distress and mistrust among the local population in its immediate aftermath despite a uniform increase in overt regime support.

The rest of the article is structured as follows. In the next section we introduce the background of the empirical study. The third section discusses the explicit and implicit measures and evaluate their construct validity. Sections 4 through 7 present the estimation framework and the results. We discuss the implications of our findings in the concluding section.

II. Background

The Purge

Our empirical analysis of preference falsification was made possible by a controversial dismissal of a high-ranking official in China. On September 24, 2006, Chen Liangyu, then Shanghai's Party chief and a Politburo member, was abruptly dismissed and placed under disciplinary investigation. The official announcement, which was broadcast nationwide on national TV in the following evening, accused Chen 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.⁵

Chen's dismissal was extensively covered by both domestic and overseas media. Chinese media toed the official line and condemned Chen's corrupt behavior.⁶ In contrast, most overseas media viewed the dismissal as indicative of a political power-struggle between Hu Jintao, then

⁵ "Decision to Investigate into Comrade Chen Liangyu's Serious Disciplinary Violations." 2006. *Xinhua News Agency*.

⁶ 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.

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 defiance of the new leadership and stubborn resistance to the central government's macroeconomic sterilization policies at a time when Hu and Premier Wen Jiabao were 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 were likely to have some believers in Shanghai after Chen's dismissal was announced. In particular, even though the power-struggle narrative was heavily censored in the Chinese media, it might still have a fairly wide audience among the local residents. Before his dismissal, Chen had enjoyed a quite high level of popularity in Shanghai, due to both his reputation as an effective administrator and his background as a local native. In a 2002 survey of public approval for mayors in 10 Chinese cities, for example, Chen, then the mayor of Shanghai, received the highest approval rating from citizens in his jurisdiction.⁷ Those who had held positive views about Chen were therefore likely to resist the idea that he was corrupt and be more receptive to the power-struggle narrative.

In Figure 1, we plot the temporal and spatial distributions of internet search activities measured by the Google search index.⁸ The search index was constructed based on the intensity of search activities within a given region (identified by IP). We focus on three key words with different underlying interpretations of the event: 陈良宇 (Chen Liangyu)—neutral, 腐败 (Corruption)—anticorruption and 上海帮 (Shanghai Gang)—power struggle. For all three

⁷ The survey is conducted by the Horizon Group, a private survey company in China.

⁸ According to a survey by China Internet Network Information Center (CNNIC), as of 2006, Google was the second most widely used search engine by Chinese internet users. See https://www.cnnic.net.cn/gywm/xwzx/rdxw/2006nrd/201207/t20120710_31486.htm

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 illustrates the spatial distribution of search interests between September and November 2006: Clearly, people in Shanghai paid much greater attention to this event than those in other part of China. The aggregate search intensity from Shanghai was 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), which is run by the China Survey and Data Center at Renmin University of China. CGSS participates in the International Social Survey Program and is one of the longest-running (first wave in 2003) and most professionally managed social surveys in China. The CGSS 2006 wave was conducted between September and November and interviewed 10151 individuals in 28 provinces. The sample in Shanghai included 400 individuals from five districts and the interviews were carried out between September 11 and November 12.⁹ Chen’s dismissal was first announced on national TV at 7:00 pm on September 25, by which time all the interviews on that day had been completed. As a result, we use September 26 as the start of the treatment period. For comparability, we exclude respondents from provinces where *all* interviews were conducted either before or after that date. This reduces our sample to 5,046 observations in 12 provinces,

⁹ The five sampled districts are Xuhui, Yangpu, Jingan, Putuo and Pudong.

with 2,638 respondents (281 in Shanghai) in the treatment group and 2363 (119 in Shanghai) in the control group. A detailed description of the survey is in the Appendix.

While Assistance from governments is necessary for large-scale social surveys in China, there is no evidence that any level of the Shanghai government was actively trying to tamper with the relevant survey responses after Chen's dismissal was made public. According to the survey organizers, government assistance comes in the form of providing sampling-related information (e.g. census and local household registration) as well as some credibility support (e.g., introducing interviewers to local residents who were suspicious about the interviewers' intention).¹⁰ Local government staff, however, were not directly involved in administering the survey or filling in the questionnaires (Bian, Li, & Cai, 2006).¹¹ Interviewers were independently recruited by the CGSS team and underwent extensive training before fielding an actual survey. The team also dispatched experienced surveyors as inspectors (*diaocha duduoyuan*) to conduct on-site monitoring of the interviewers' behavior and to check for irregularities in completed surveys.

III. Gauging Preference Falsification: Strategies and Measures

To assess public responses to Chen's dismissal and, more importantly, distinguish preference falsification from genuine conviction, we need 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

¹⁰ The presence of local government staff during the interview is indeed one of the main forms of political pressure that many respondents may face.

¹¹ Since the content of the survey would not be published after at least two years and the questionnaire contained numerous items, government staff also had few *incentives* to tamper with the results.

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 insincere responses from sincere ones.

The natural coincidence between the announcement of Chen's dismissal and the survey allows us to 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 nonetheless indirectly related to 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, & Szechrest, 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 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 the two sets of measures, on the other hand, can be evidence for preference falsification.¹²

¹² We formalize this idea with a simple model in the Online Appendix.

The selection of explicit and implicit measures is guided by both our substantive knowledge about the nature of the event and the availability of questionnaire items in the CGSS. Our prior knowledge about the event leads us to expect that the event will not only affect respondents' expression of political support but also elicit polarized responses from supporters and opponents of the dismissal in domains such as general emotional well-being, views on the legal institutions, and perceptions of the severity of official corruption. Guided by these expectations, we look specifically for questions that belong to these domains from the survey questionnaire. As the survey was primarily designed and fielded by a team of sociologists, there are a limited number of questions on political attitudes that can be used for our research purpose. After ruling out obviously unrelated questions,¹³ we were able to pick about 5 to 6 candidate questions for each type of measures.¹⁴ Additional analyses of the basic patterns of response allowed us to further drop questions that either have too many missing responses or demonstrate a pattern of response that is clearly different from the rest of the questions.¹⁵ This leaves us with three questions for each type of measure. All selection of questions was conducted prior to the data analysis.

We choose the following three questions as the explicit measures for political 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.)

¹³ For example, a large share of CGSS's questions on political attitudes is on nationalism, which is not relevant for our research purpose here.

¹⁴ For explicit measures, for example, we focus on Section E.47 of the survey questionnaire and pick all the questions that contain the words “政府 (government)” or “民主 (democracy)”.

¹⁵ More specifically, for each type of measure, we conduct principal component analysis on all candidate questions and then drop the item with the smallest loading on the first principal component. We repeat this exercise for 3-4 times until the all remaining questions have sufficiently high loadings (>0.4) on the first principal component.

- **[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.¹⁶ 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 is perceived pressure for displaying political support. If this was the case, then all respondents would give more affirmative responses to these

¹⁶ 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 higher than the actual level of regime 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 happy)

- **[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 ...) ¹⁷

We reverse the scale for the second question (making “strongly disagree” to 4 and “strongly agree” to 1) to make it 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 be directly pertinent to one’s overall regime support. The second and third questions only ask for assessments of certain parts of the system (courts and officials), but not about the regime as a

¹⁷ Other choices include: (4) managers of state-owned enterprises, (5) managers of collective enterprises, (6) private investors, (7) foreign investors, (8) self-employed, (9) managers in foreign and private firms, (10) intellectuals, and (11) others.

whole (in follow-up experiments, we will show that there is no consensus on which answers are more politically desirable). In the specific 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 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),¹⁸ and potentially even have their views about the judiciary improved upon seeing the law enforcement actively cracking down on corruption.

In contrast, those who sympathized with Chen and were skeptical about the official narrative are likely to give the exactly opposite responses to the implicit measures. Their happiness is likely to decline for three reasons. First, they might lament the loss of a competent native leader who was willing to stand up for local interests and in turn be displeased with the central leaders’ political machinations. Second, they might also be worried that the purge of Chen would be followed by other political and economic measures against Shanghai, which would negatively impact their own socioeconomic interests. At a deeper level, those who interpreted the dismissal as a purge might also draw parallel with other unpleasant political experience in the past. In particular, the Cultural Revolution, which had disrupted lives of millions of ordinary citizens for

¹⁸ Theoretically, it is possible that those who viewed the case as a genuine corruption case also chose other categories, such as “private investors” in responding to this question, as these groups also frequently appear in corruption narratives. However, a survey of both domestic and overseas reports on Chen’s dismissal during the period of interest suggested that there was relatively little mention of the business actors involved. The same DID regression using a binary variable for whether private investor was chosen as the greatest beneficiary yields negative and insignificant coefficient (point estimate=-.022, p value=0.667). We also conducted a robustness check in which the beneficiary group dummy is coded as 1 if the respondent chooses either government officials *or* the private investors, and the results remain virtually unchanged. The detailed results are available upon request.

more than a decade (and especially in Shanghai), had similarly started with a series of purges of high-ranking officials (MacFarquhar & Schoenhals, 2006).

Relatedly, since the Party's disciplinary investigations would inevitably be followed by a guilty conviction in the court (Manion, 2004), the purge of Chen was also likely to remind Chen's sympathizers about the court's lack of independence when they were asked about their views toward the judiciary. Moreover, since they did not consider the anticorruption case against Chen as legitimate, they were unlikely to treat it as new information about the state of official corruption.¹⁹ As a result, their responses to the third question would not change much.

The predicted changes for supporters and opponents of the purge are summarized in Table 1. In the Online Appendix, we test the validity of our predictions in a follow-up online experiment in which 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.

[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

¹⁹ In fact, some even considered Chen as "exceptionally clean" after the authority revealed that his case only involved the misuse of a couple of million yuan. See, for example, <http://shihuaningvip.blog.sohu.com/83609616.html> (in Chinese)

two groups of respondents to assess the perceived sensitivity of a list of questions, including both our explicit and implicit measures.²⁰ 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. The first column on the left shows that all three explicit measures are rated as more sensitive than the implicit measures. The second panel plots the average difference in responses between the “most people’s answer” group and the “your real answer” group. Again, there are significant differences between the two groups for all explicit measures but no such difference exists 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 should 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 missing response are on average higher for the explicit measures than the 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]**.

²⁰ 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.

IV. Empirical Strategy

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:

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

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,²¹ 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 potential imbalance, we use entropy balancing (EB), which is a re-weighting technique developed by Hainmueller (2011), to adjust the distributions of key covariate in the treatment and the control groups. The EB algorithm calculates a weighting scheme that achieves almost exact balance on the first and higher moments of targeted covariates specified by the researcher. We choose to balance on the following demographic variables: *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 dataset.

V. Results

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 5th order polynomial lines for all the

²¹ 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.

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 the national trend whereas actual support dropped significantly below it.

[Figure 3 about here]

Next, we use regression analyses to assess the statistical significance of these visual patterns. Table 2 presents the coefficient 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 three original questions from which the synthetic measures are constructed. The last four columns present the estimates using the two synthetic measures created by 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 1st. The exclusion of the post-Nov 1 interviews is motivated by the consideration that the salience of Chen's downfall had probably dissipated by then.

[Table 2 about here]

The regression results are consistent with our impressions from the visual analysis: The announcement of Chen's dismissal had substantial impacts on both the expressed and the actual attitudes of Shanghai respondents, but in *opposite* directions. For the explicit measures, the coefficient estimates are positive and significant for all but one original questions 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. For each of the original questions, 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 the aggregation method and the sample choice, the coefficient estimates are consistently positive and significant and the magnitudes are as large as 35% of a standard deviation.

Turning to the implicit measures, however, the estimates suggest that Shanghai residents were indeed 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 (+10% SD and 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 there was a sharp increase in expressed support following the announcement of Chen's dismissal. This increase, however, 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.

VI. Testing Alternative Explanations

In this section, we conduct additional tests to evaluate the robustness of our main results and to address several key alternative explanations. First, one might raise the concern 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 had a low opinion about Chen’s dismissal did 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, Shanghai sample only) separately for the treatment and control periods. We find a strong negative association between the residuals of the two measures in the treatment period ($\rho_{\text{Treatment}} = -0.178$, $p < 0.01$),²² 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 pressure to falsify one’s preferences is stronger.

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 could exogenously affect the degree of political pressure a respondent faced. One place to look for such an

²² To see if the strong correlation in the post-treatment was driven by a few outliers, we also obtain alternative estimates using robust regression, which is a technique that specifically addresses the issue of influential data points. The point estimate is -0.158 with a standard error of 0.062, which is still significant at 5% level.

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 should directly alter the respondent's genuine support for the regime, it is highly plausible that respondents would have greater incentives to express a higher level of political support when being interviewed by an older person, who was more likely to be mistaken for a state agent than a young student.²³

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 ($\beta_{\text{Control}} = -0.019$, $p < 0.19$).²⁴ This finding provides further support to the claim that individual-level consideration about political

²³ The interviewers for the CGSS are typically recruited from two major sources: students (college or high school seniors) and laid-off workers (*xiagang gongren*) (Bian et al., 2006).

²⁴ We also obtain alternative estimates using robust regression. The estimated correlation for the post-treatment period is 0.019 (se=0.009), which is statistically significant at 5% level. 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.

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.²⁵ 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.²⁶ 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 announcement of Chen's purge.²⁷ The *difference* between the two estimates (highlighted by the red dashed line) is unambiguously maximized on September 26.

²⁵ 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.

²⁶ 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.

²⁷ 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.

[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 in which 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]

Finally, an alternative interpretation for the divergence is that while the Shanghai respondents were unhappy about losing their favorite leader, they nonetheless became genuinely supportive of the political authority because forceful action against an unruly local official signaled that the new national leadership was strong. While it is impossible to directly assess how respondents were thinking at that time, we can test two implications that follow from this interpretation. First, if the increase in trust was due to the perception of a stronger political authority, we should also expect such perception to make the respondents more willing to delegate socioeconomic responsibilities to the government. Second, perceptions of a stronger political leadership might

also lead respondents to become more optimistic about their own socioeconomic well-being in the future, in spite of the short-term frustrations with losing a beloved local leader.

To test the first implication, we utilize questions in the CGSS that ask whether the respondent think a certain industry (a total of ten were listed)²⁸ should be run by the state, the private sector, or both. For each respondent, we calculate the total number of industries that he/she thought should be run by the state as the dependent variable for the first test. To test the second implication, we use a question that asks the respondent's view about his family economic condition in the next three year.²⁹ We construct the second dependent variable as a dummy that takes the value of 1 if the respondent thinks the condition is going to get better and 0 otherwise.

Table 4 displays the results from regressing these two variables on the baseline specification. We find that there was some decline the number of industries that Shanghai respondents thought the state should run following the purge, but the change was not statistically significant. Moreover, there was also no significant increase in respondents' confidence in future well-being. These results suggest that at least changes in observed attitudes do not seem to support the view that the increase in trust was driven by the perception of a stronger national leadership.

[Table 4 about here]

VII. Subgroup Analysis: Who Are the Liars?

²⁸ These are: postal services, electricity, banking, steelmaking, railway, hospital, museum/parks, shopping mall, university, and agricultural procurement.

²⁹ Specifically, the question asks "What kinds of change would you expect for your family's economic condition in the next 3 years". Choices are (1) getting better, (2) no change, and (3) getting worse.

We have so far provided evidence that the marked divergence between expressed and actual support in the aftermath of purge of Chen Liangyu was most consistent with the interpretation of preference falsification. In this section, we explore the variations in the extent of falsification across different subgroups.

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, a key determinant for developing unorthodox political beliefs is access to *alternative sources of information* about politics that helps the recipients resist and argue against the regime's dominant 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. Punishment on undesirable political expressions can be imposed directly by the state in the form of political or financial sanctions, but can also take the form of harassment or social ostracization by ordinary people who wish to signal their political loyalty. In both cases, we expect that those who had more to lose from these potential sanctions would be more likely to express supportive attitudes in their responses.

With these conjectures in mind, we examine a wide range of subgroups, including political affiliations, sector of employment, basic demographic groups, and social attitudes. For each

subgroup, we run the same regression model as Model 6 in Table 3 using expressed support, actual support, and their difference (the falsification index) as the dependent variables, respectively. The results, displayed in Figure 6, are ordered by the magnitude of the treatment effects on the falsification index.

[Figure 6 about here]

It is evident that while the difference between expressed and actual support expanded for most of the subgroups in Shanghai during the aftermath of the purge, there were marked variations in the *degree* of falsification across different subgroups. Starting from the top, we find that falsification is most intense among (1) wealthy individuals (the top 20% of the local income percentile),³⁰ (2) those with college-level education or above,³¹ (3) frequent internet users,³² and (4) public sector employees. Memberships in the first three groups are highly correlated.³³ Together, they represent the segment of Chinese society that is well informed, intelligent, and with relatively high socioeconomic status.

Based on our conjectures on information access, it is not surprising that these groups distrusted 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 seen their personal wealth improved

³⁰ 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.

³¹ 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.

³² Frequent internet users are defined as those who use internet on a *daily* basis.

³³ 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.

under his administration. Frequent internet users had better access to overseas coverage that interpreted Chen's dismissal as the result of a power struggle. However, it is somewhat counterintuitive to observe that these groups were also among those that tried hardest to hide their actual opinion by reporting a high level of support for the regime after the purge. In particular, the behaviors of the internet users were in sharp contrast to the critical and activist style of they took in the on-line communities (King, Pan, & Roberts, 2013; Yang, 2013). Here, our second conjecture can provide an explanation: These groups, by virtue of their better economic conditions, typically had more to lose from potential political 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 some 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.

Public sector employees displayed the fourth highest degree of falsification among all subgroups. This, again, is consistent with our theoretical conjecture: 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 give them strong incentives to hide their real opinions and demonstrate loyalty at politically sensitive moments.

Compared to the public sector employees, those who work in the private are usually less dependent on the state and thus have fewer incentives to feign a supportive attitude. This distinction is reflected in the subgroup results: Like their public-sector counterparts, private

sector employees also reported a significant decline (of even larger magnitude) in actual support after the purge. However, unlike the former, there was no simultaneous increase in expressed support.³⁴

Finally, we also find interesting variations in falsification across different age cohorts. In particular, the degree of falsification was most intense among those between the ages of 50 to 60, a cohort that had spent most of their formative periods in the Cultural Revolution (1966-1976). Living through a tumultuous period during which the personal cult of Mao was at its peak had probably taught people in this cohort the importance of toeing to the Party line, making them particularly prone to feigning supportive political attitudes when the political atmosphere was tightened.

VIII. Concluding Remarks

Much of the knowledge creation in social sciences today still relies on self-reported data, which may suffer from various forms of biases, including respondents' deliberate misreporting. In this study we employ a novel method that combines conventional surveys with natural experiments to study how the Shanghai public reacted to the controversial purge of a popular local leader, and how such reactions varied across different social groups. Our analysis reveals a marked divergence between expressed and actual regime support among the Shanghai respondents in the aftermath of the purge. We interpret this divergence as evidence of systemic preference falsification after ruling out a number of alternative explanations. We also find that the variations in the degree of preference falsification across subgroups are best predicted by the

³⁴ Another group of people often considered to be political insiders are CCP members. Their estimates, however, are statistically insignificant in both dimensions. 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.

combination of information exposure and vulnerability to political sanctions. In particular, contrary to the conventional view that citizens who are wealthier, better educated and more informed are more likely to voice out criticisms about the authority, our analyses show that they are in fact among the groups that tried to hide their true beliefs to the greatest extent when the potential cost of expressing dissent increases.

This study has several implications for research on comparative mass behavior in general and public opinions in authoritarian regimes in particular. Methodologically, the design-based approach employed in this study offers an alternative strategy for detecting preference falsification or social desirability in settings where conducting controlled experiments is difficult or infeasible. This method is not only useful for researchers of authoritarian regimes to circumvent political obstacles on conducting in-depth opinion surveys or large-scale controlled experiments, but can also be deployed by researchers of democracies to study how the public respond to sensitive events using existing survey data. One may, for example, apply the same method to study how recent significant public events have altered public opinions on race, religion, and immigration in the United States and some European countries by comparing responses to questions on explicit and implicit attitudinal measures both before and after those events. As opinion polls are regularly being conducted in many democratic countries, there will be no dearth of surveys to work with. Future research may further improve this method by developing questions with better controlled content and embed them in regular surveys in advance. In this way, the design will allow researchers to combine the strong internal validity of controlled experiments with the external validity from using real-life events as treatment.³⁵

³⁵ In a design that shares the spirit with ours, Schmidt and Nosek (2010), for example, have conducted lab experiments on a large sample collected before, during, and after the Obama's presidential campaign to evaluate changes in explicit and implicit racial attitudes of the subjects.

In terms of substantive findings, although our analysis focuses on attitudinal changes around a single event, the central insight that preference falsification remains widespread in dictatorships during politically sensitive moments is likely to hold in other contexts as well. Although not of daily occurrences, political tension and uncertainty are by no means rare for those living in 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 many more would be watching their words during Stalin's Great Purges, Mao's Anti-Rightist campaigns, Great Leap Forward, and Cultural Revolution, or after crackdown of mass social movements, such as the Tiananmen protest.

In addition to these historical parallels, the findings also bear relevance for our understanding of some major on-going events in China. The recent anticorruption campaign launched by Xi Jinping, for example, had similarly led to the downfall of many locally popular figures,³⁶ and the enforcement of strict disciplinary measures have had even wide impact on the livelihood of a large number of cadres. So far, Xi's campaign, like the one we documented in this study, have been greeted with almost uniform support and enthusiasm from the public.³⁷ While we do not deny that there might be much sincerity in the support expressed, given the unprecedented scope and intensity of the most recent campaign, the overall degree of support is most likely overstated, especially when the central leadership has not only made clear their stance on these matters but also taken a heavy hand in suppressing dissident views (CHRD, 2015).

³⁶ For example, Bo Xilai, who was among the first to fall after Xi took power, enjoyed high level of popularity in Chongqing. A number of local officials who were recently sacked also had local reputation as the "competent men". See <http://cpc.people.com.cn/pinglun/n/2014/1103/c241220-25964896.html> (in Chinese).

³⁷ According to a recent poll by the Pew Center, 63% of the respondents think that the problem of corruption will be less severe in five years (see <http://www.pewglobal.org/2015/09/24/corruption-pollution-inequality-are-top-concerns-in-china/>). Polls conducted by the Chinese authority similarly suggest that over 73.7% of the mass and about 90% of the cadres think the authority suggest that they supported the state's anticorruption campaign. See <http://opinion.people.com.cn/n/2014/0103/c1003-24011495.html> (in Chinese)

The finding that the rich, educated, and internet-savvy citizens tend to falsify their opinions to the greatest extent in face of political pressure also bears implications for our understanding of the prospect and patterns of political change in nondemocratic systems. While the prevailing theories of regime transition suggest that economic growth, rising levels of education, and better access to information all contribute to the likelihood of democratization at the aggregate level (Epstein, Bates, Goldstone, Kristensen, & O'Halloran, 2006; Lipset, 1959), it does not imply that the wealthy, educated and informed *individuals* will necessarily be the pioneering agents in such transitions. Quite to the contrary, our analysis suggests that these groups may be the least likely to openly oppose the regime, precisely because they are sophisticated enough to appreciate the consequences of such actions and have much to lose when sanctions are imposed. The first call for change, therefore, may have to come from somewhere else—most likely those who are sufficiently independent from the state or have relatively little to lose from potential political punishments.

References

- Arendt, Hannah. (1972). *Crises of the Republic: Lying in Politics; Civil Disobedience; On Violence; Thoughts on Politics and Revolution*: Mariner Books.
- 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
- Bian, Yanjie, Li, Lulu, & Cai, He. (2006). *Shehui Diaocha Fangfa yu Jishu: Zhongguo Shijian (The Method and Practice of Social Survey: Experience from China)*: Shehui Kexue Wenxian Chubanshe.
- 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.
- Bradburn, Norman.M., Sudman, Seymour., & Blair, Edward. (1979). *Improving interview method and questionnaire design*: Jossey-Bass.
- Chen, Jie. (2004). *Popular Political Support in Urban China*: Stanford, California: Stanford University Press.
- CHRD. (2015). Silencing the Messenger: 2014 Annual Report on the Situation of Human Rights Defenders in China: Chinese Human Rights Defenders.
- Cox, Daniel, Jones, Robert P., & Juhem, Navarro-Rivera. (2014). I Know What You Did Last Sunday: Measuring Social Desirability Bias in Self-Reported Religious Behavior, Belief, and Identity: Public Religion Research Institute.
- 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
- Epstein, David L., Bates, Robert, Goldstone, Jack, Kristensen, Ida, & O'Halloran, Sharyn. (2006). Democratic Transitions. *American Journal of Political Science*, 50(3), 551-569. doi: 10.1111/j.1540-5907.2006.00201.x
- Farrell, Henry. (2012). The Consequences of the Internet for Politics. *Annual Review of Political Science*, 15(1), 35-52. doi: doi:10.1146/annurev-polisci-030810-110815
- Fazio, Russell H., & Olson, Michael A. (2003). Implicit Measures in Social Cognition Research: Their Meaning and Use. *Annual Review of Psychology*, 54, 297-327.
- 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
- 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
- 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. (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>
- Kurzban, Charles. (2009). *The Unthinkable Revolution in Iran*: Harvard University Press.
- Lee, Zoe, & Woodliffe, Lucy. (2010). Donor Misreporting: Conceptualizing Social Desirability Bias in Giving Surveys. *Voluntas: International Journal of Voluntary and Nonprofit Organizations*, 21(4), 569-587.
- 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.
- Lipset, Seymour Martin. (1959). Some Social Requisites of Democracy: Economic Development and Political Legitimacy. *American Political Science Review*, 53, 69-105.
- 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).
- MacFarquhar, Roderick., & Schoenhals, Michael. (2006). *Mao's Last Revolution*: Belknap Press of Harvard University Press.
- 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
- Paulhus, Delroy L. (1984). Two-Component Models of Socially Desirable Responding. *Journal of Personality and Social Psychology*, 46(3), 598-609.
- Payne, J. Gregory. (2010). The Bradley Effect: Mediated Reality of Race and Politics in the 2008 U.S. Presidential Election. *American Behavioral Scientist*, 54(4), 417-435. doi: 10.1177/0002764210381713
- 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.
- Schmidt, Kathleen, & Nosek, Brian A. (2010). Implicit (and explicit) racial attitudes barely changed during Barack Obama's presidential campaign and early presidency. *Journal of Experimental Social Psychology*, 46(2), 308-314. doi: <http://dx.doi.org/10.1016/j.jesp.2009.12.003>
- 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
- 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
- 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.
- 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.
- 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.

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)	+	+	+
Opponents (Sympathized with Chen)	-	-	No significant 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 (0.204)	-0.059 (0.149)	0.001 (0.139)	-0.066 (0.131)	-0.034 (0.120)	0.120 (0.269)
Treatment x Shanghai				0.068 (0.207)	0.065 (0.195)	-0.090 (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.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

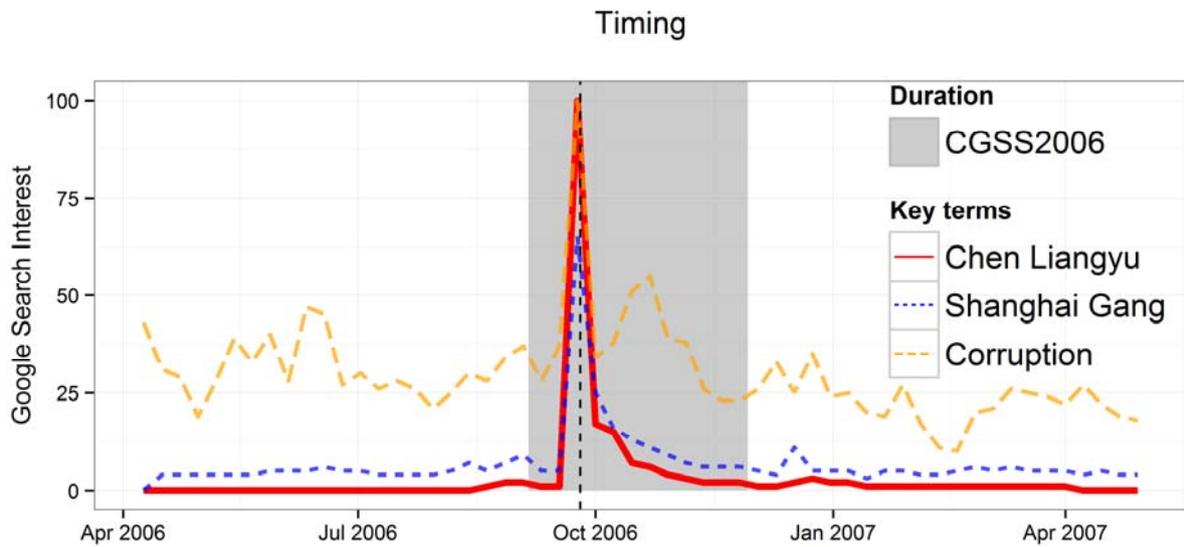
Table 4 Alternative Explanation: Stronger Leadership

	# of Industries State can Run		Life Better in 3 Years	
	(1) Full	(2) Short	(3) Full	(4) Short
Treatment x Shanghai	-0.270 (0.289)	-0.248 (0.290)	0.105 (0.121)	0.105 (0.121)
District FE	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes
Observations	5046	5000	5046	5000

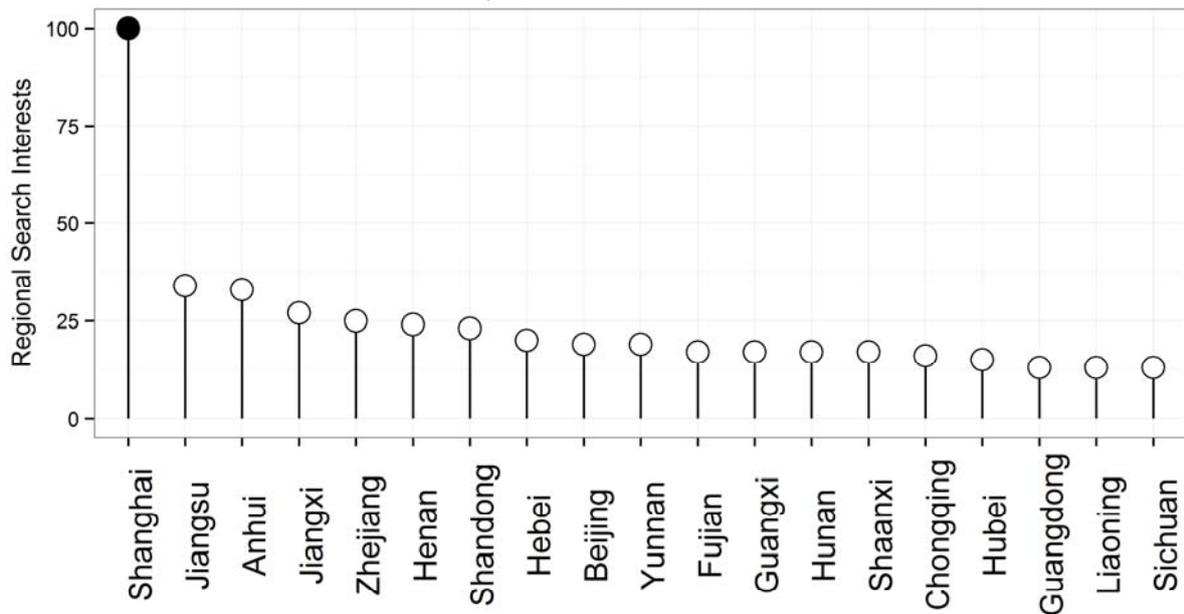
Notes: Analyses are based on 5 multiply imputed datasets. The demographic controls include *age, gender, sector of employment, Party membership, college degree, ethnicity, marital status, household registration status (hukou), employment status and military experience*. Robust standard errors clustered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

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

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

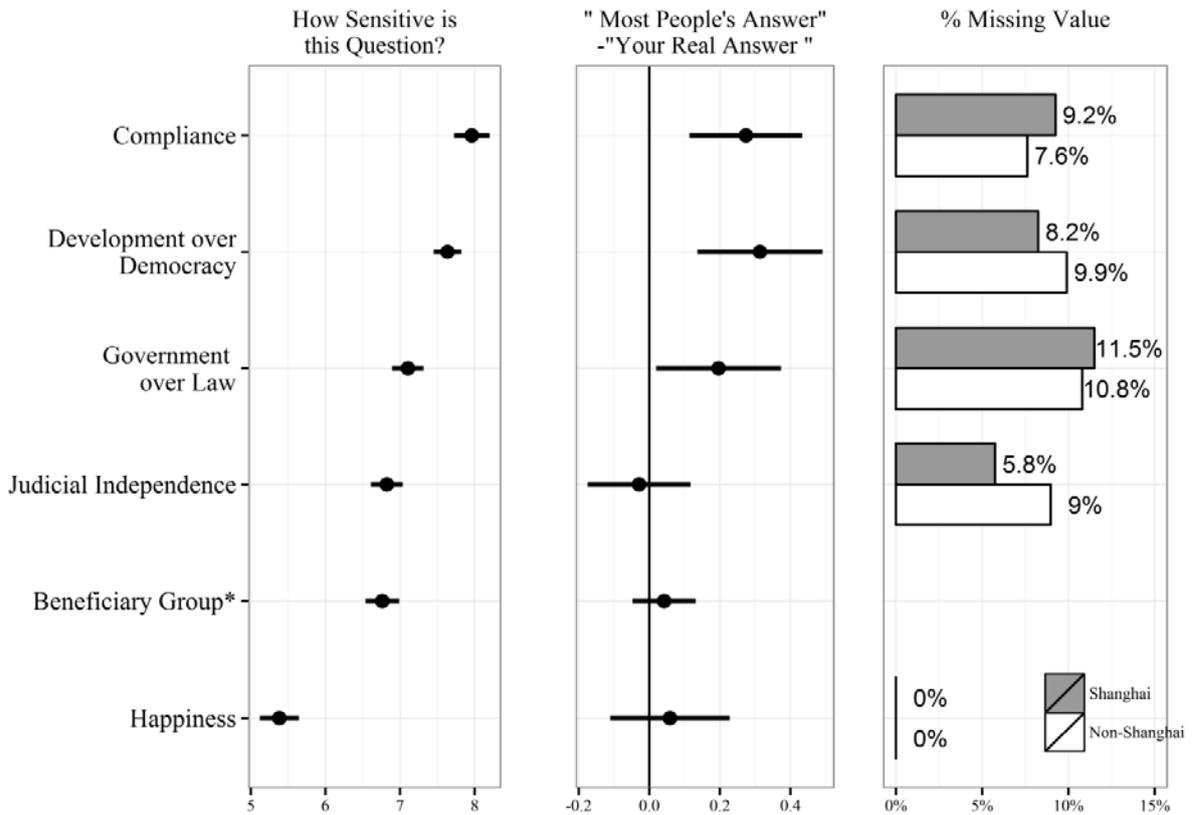


Regional Search Interest for Keyword "Chen Liangyu":
September~November 2006



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.

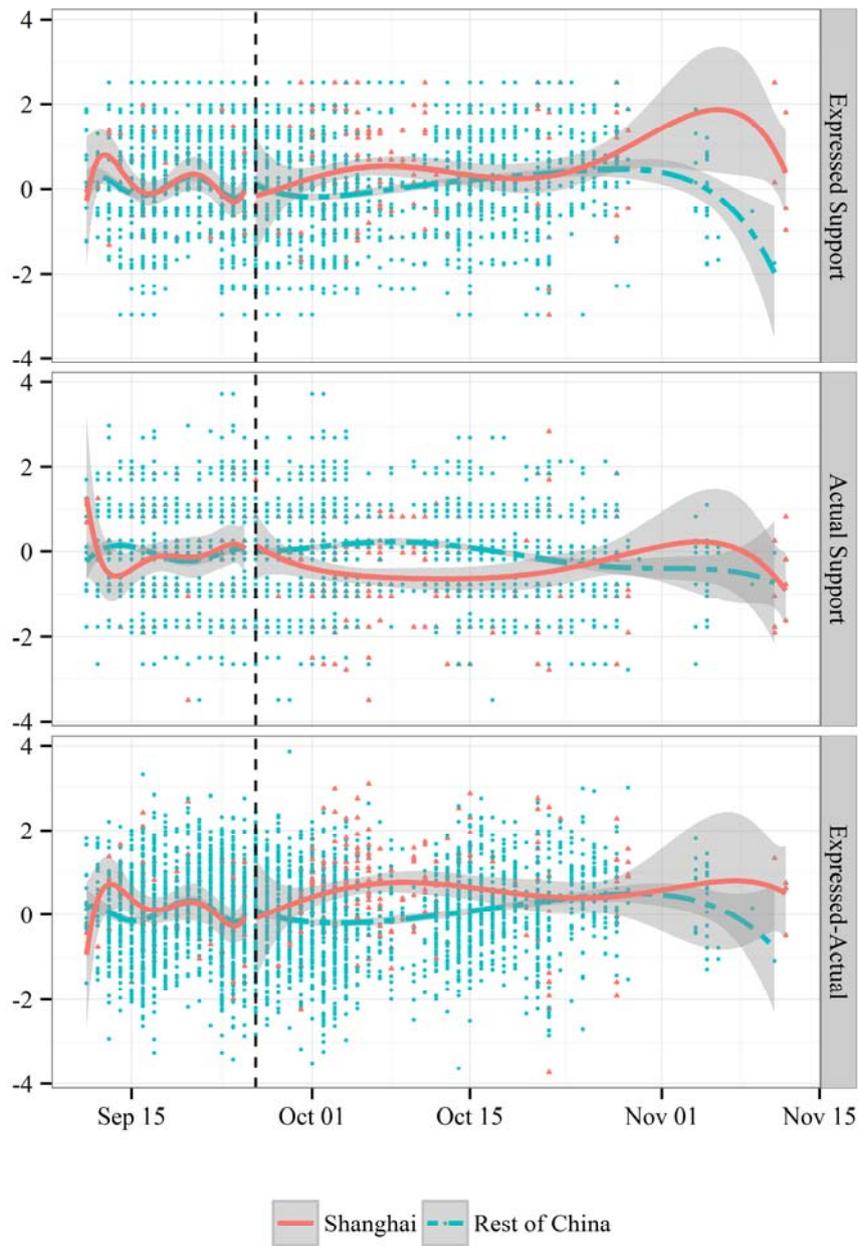
Figure 2: Evaluating the Relative Sensitivity of the Explicit and Implicit Measures



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.

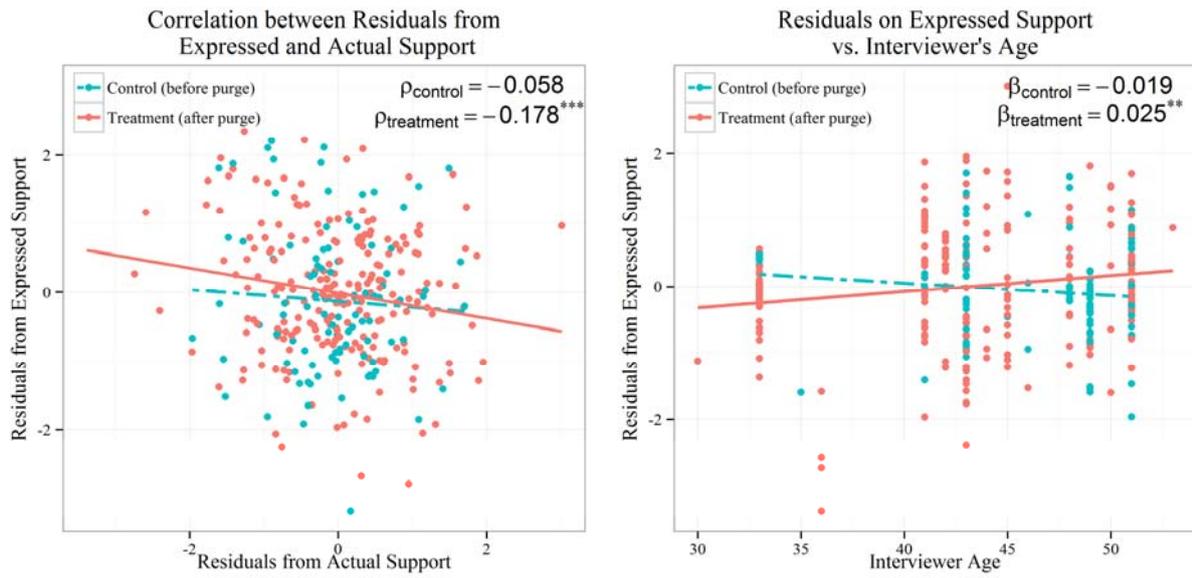
* Since this question does not allow no-response, we do not plot its missing rate on the right panel.

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.

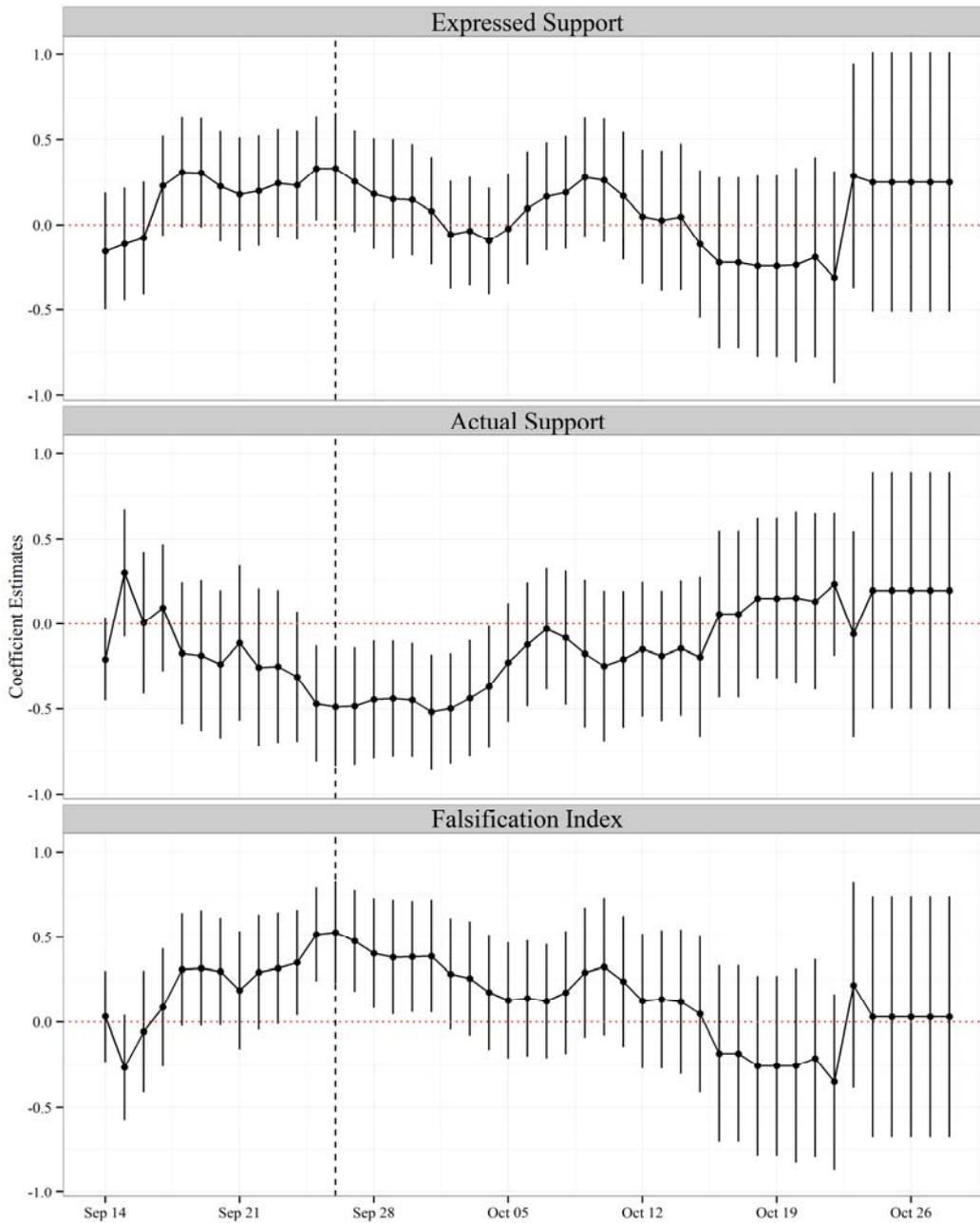
Figure 4: Analysis of Residuals



Note: In each panel, two separate linear 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.

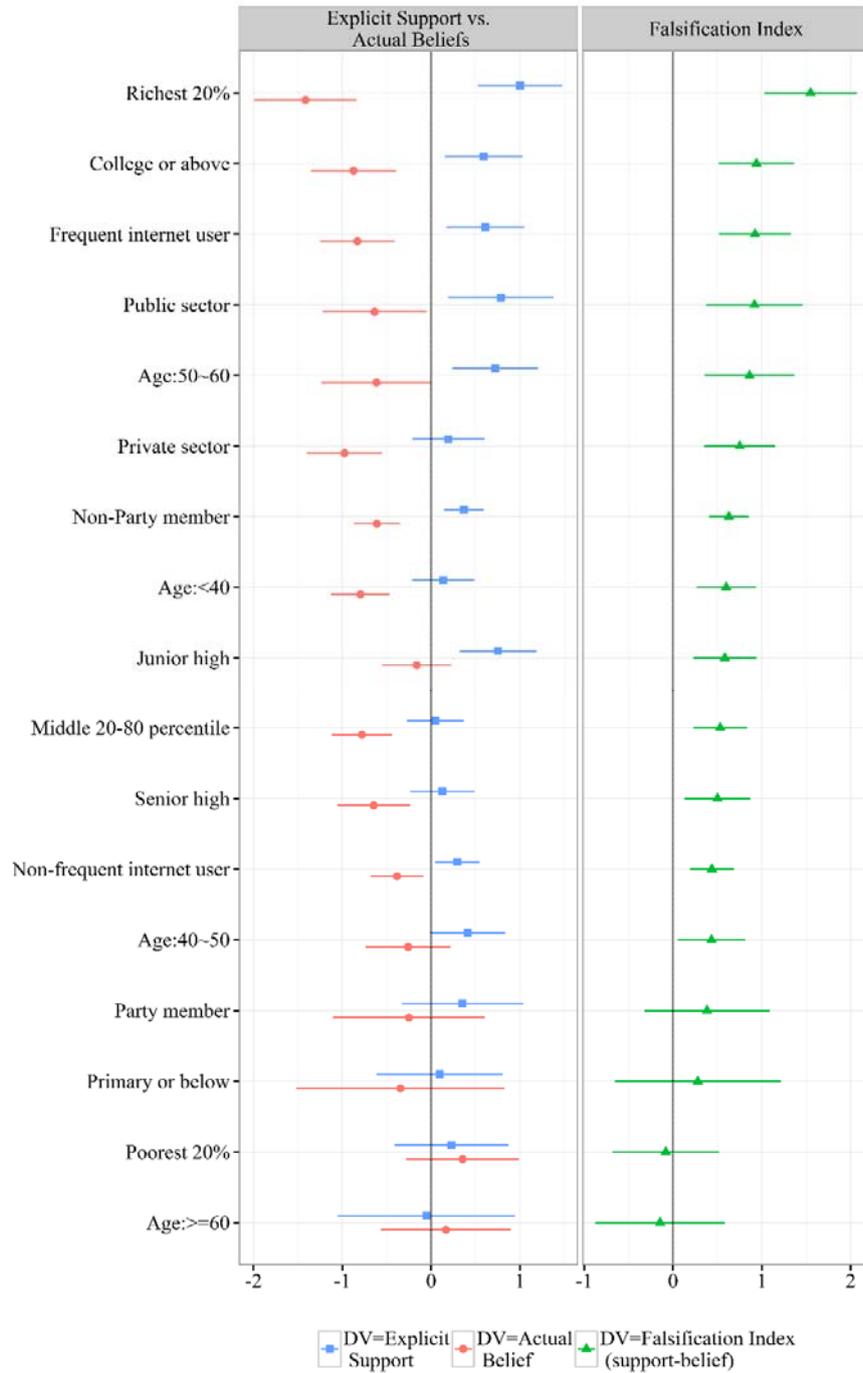
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure 5 Testing Different Cutoff Dates



Note: Each dot represents the DID estimate for the Shanghai sample using different cutoff dates (indicated on the x axis) for expressed and actual support as well as their difference (falsification index). The vertical dashed line marks the actual cutoff on the date of September 26 2006. Vertical bars indicate the 90% confidence intervals.

Figure 6 Heterogeneous Effects of Purge on Expressed and Actual Support



Notes: Each circle represents the treatment effects (Treatment x Shanghai) on a given dependent variable (indicated on the top). Horizontal bars indicate the 90% confidence intervals. Estimates are based on the same specification as Model 6 of Table 2.