SHORT ARTICLE

What Makes Anticorruption Punishment Popular? Individual-Level Evidence from China

Lily L. Tsai, Massachusetts Institute of Technology
Minh Trinh, Massachusetts Institute of Technology
Shiyao Liu, New York University Abu Dhabi

How does punishment of corruption help to build public support in authoritarian regimes? We outline two primary mechanisms. Instrumentally, the ability to pursue anticorruption initiatives to the end signals government capacity. Deontologically, anticorruption punishment signals moral commitments. Through a novel experiment design for mediation analysis embedded in a series of conjoint experiments conducted in China, we find individual-level evidence to support both mechanisms. Specifically, we find that Chinese citizens positively view local government officials who punish their corrupt subordinates and that this positive view arises out of the perception that these officials are both competent in their jobs and morally committed to citizens' value. The preference for anticorruption punishment is substantial compared to other sources of public support in authoritarian regimes—economic performance, welfare provision, and institutions for political participation—suggesting that it could become a popular strategy among autocrats.

In recent years, governments in nondemocratic and hybrid regimes have increasingly punished corruption. While some anticorruption campaigns primarily consist of messaging, punishment, often severe, has become salient in places like Rwanda, Tanzania, Iran, and China, and such initiatives have been popular among autocrats for decades. Although whether punishment actually reduces corruption remains debatable, it is increasingly plausible that authoritarian regimes sometimes choose this strategy for its popularity with citizens.

Why is anticorruption punishment effective at bolstering public support for authoritarian regimes? We propose an individual-level explanation, arguing that anticorruption punishment can improve citizens' perceptions of the officials who carry it out. This effect operates in part through a deontological channel: citizens positively view officials who punish their corrupt subordinates because they think these officials are morally aligned with citizens' values. This channel helps us understand why leaders in authoritarian regimes might be genuinely popular and may explain how anticorruption punishment increases support even when citizens have no evidence that corruption itself is decreasing. It also suggests that leaders who convince the public that they share the public's moral values can garner their support, even in places like China, where citizen perceptions of government competence are already high.

We demonstrate this argument through a series of novel conjoint experiments administered in China, which find that Chinese citizens prefer local government officials who engage in top-down punishment of corruption by their subordinates. Using an experimental design to conduct mediation analysis, we find evidence for both an instrumental and a deontological basis for this preference.

Lily L. Tsai (l_tsai@mit.edu) is the Ford Professor of Political Science at the Massachusetts Institute of Technology, Cambridge, MA 02139. Minh Trinh (mdtrinh@mit.edu) is a PhD candidate in the Department of Political Science at the Massachusetts Institute of Technology, Cambridge, MA 02139. Shiyao Liu (ssliu@nyu.edu) is a postdoctoral associate at New York University Abu Dhabi, Abu Dhabi, United Arab Emirates. Author names are listed in reverse alphabetical order.

All studies were conducted in compliance with relevant laws and were approved or deemed exempt by the appropriate institutional review board. Data and supporting materials necessary to reproduce the numerical results in the article are available in the *JOP* Dataverse (https://dataverse.harvard.edu/dataverse/jop). An online appendix with supplementary material is available at https://doi.org/10.1086/715252.

Published online October 25, 2021.

 ooo

ANTICORRUPTION PUNISHMENT AND PUBLIC OPINION

Since the 1990s, scholars of the Soviet Union have argued that the real purpose of anticorruption initiatives was to boost public support instead of improving government performance (Holmes 1993, 204). More recently, anticorruption campaigns such as those being carried out in China under Xi Jinping are widely seen as a strategy to revitalize popular support (Xi, Yao, and Zhang 2018), even when they may have other negative impacts (Wang 2019). The utility of anticorruption initiatives seems clear, given the broad empirical literature documenting the negative link between perceived corruption and public support in electoral democracies (e.g., Ferraz and Finan 2008).[1] However, even when citizens lack hard evidence that punishment reduces corruption, it may still increase public support.

Existing literature suggests two main explanations for why anticorruption punishment might improve public support for authoritarian governments. First, punishment may signal capability—the government is competent enough to enforce its own rules. Citizens may thus see authorities who punish corruption as more capable of ensuring that their subordinates follow policies and better at deterring future corruption, all other things being equal (Levi and Stoker 2000). They may also infer that the government is generally capable of achieving its goals, including delivering on promised governance outcomes.

Apart from this instrumental mechanism, punishment may send signals about the moral commitments and benevolent intentions of the government. Because citizens see corruption negatively, tackling corruption signals that the government shares their normative values. In authoritarian regimes, anticorruption punishment makes this signal credible, because it shows that autocrats are committed enough to take the costly step of sanctioning other regime insiders. Observing the punishment of some officials, citizens may believe that the system as a whole is otherwise good, despite some bad officials (Bies 1987; Festinger 1950). This deontological logic may explain how anticorruption punishment improves public opinion about the regime, even if citizens have no evidence that corruption itself is decreasing.

These two explanations for anticorruption punishment's effect on public opinion are not mutually exclusive. However, for either of them to be valid, we must observe evidence indicating that individual citizens are indeed making instrumental or deontological inferences when evaluating government authorities upon observing these officials' anticorruption behavior. Specifically, all else being equal, individual citizens prefer officials who punish their lower-level subordinates for corruption over those who do not. If it were to influence support for the highest authority, this preference should manifest when citizens evaluate local officials with whom they interact the most.[2] It must also be substantively important, meaning that it should measure up to other considerations that have been found to drive public opinion in authoritarian regimes. Critically, it must operate through both competence evaluations and moral character evaluations: citizens who perceive information about an official's anticorruption punishment must make inferences about this official's level of competence and moral quality, and these inferences must in turn influence their overall evaluation of this official.

EVIDENCE FROM CHINA

We embed a set of novel conjoint experiments in two face-to-face surveys implemented in China. Given possible preference falsification (Jiang and Yang 2016), conjoint experiments are especially useful (Hainmueller, Hopkins, and Yamamoto 2013). In study 1, which was conducted in rural counties of the Fujian province ($n = 320$), we presented respondents with profiles of two hypothetical township party secretaries and asked them who they would prefer to have serving in their own township. Even though Chinese citizens do not vote for township-level officials (Manion 2006), our qualitative interviews indicated that they interpreted this question to ask about their attitudes, not about whether they could literally choose whom they wanted in office. We explicitly avoided electoral language, choosing instead to ask respondents about their comparative feelings toward the two profiles.

As the main experimental treatment, each profile is randomly chosen to show a track record of having (or not having) punished corruption by village officials below them.[3] This information is presented as a conjoint attribute, which we coded as Punishment. To benchmark the importance of anticorruption punishment, we presented three additional randomized attributes, corresponding to other considerations that scholars argue are essential in shaping citizen evaluations of government authorities: the Growth attribute signals whether the officials have promoted local economic growth (e.g., Chen 2004; Dickson 2016); the Welfare attribute highlights whether

1. Wang and Dickson (2019) show that investigating corrupt officials may reduce regime support if it reveals previously undetected acts of corruption, but they do not explore the effect of punishing these officials.

2. Li (2016) shows that central authorities enjoy higher baseline levels of public support than local ones.

3. In our pilot research, we found that citizens do not have enough information to precisely know how much actual corruption has changed after each anticorruption initiative. To match these real-life conditions, we designed the experiment to evaluate the impact of anticorruption punishment in the absence of information about its effect on the frequency of corruption.

they have implemented minimum income guarantees, the most basic and well-known form of social welfare in China (e.g., Whyte 2010); and the Election attribute shows whether they have guaranteed meaningful village elections, Chinese citizens' primary channel for regular political participation (e.g., Dickson 2016; Manion 2006). Following Meyer and Rosenzweig (2016), our enumerators introduced the attributes using predefined statements and cartoon images that effectively convey their meanings.[4]

For this study, we asked each respondent to do six different rounds of the conjoint experiment, with the profiles being rerandomized each time as in a fully crossed design. Respondents thus choose between two random "baskets" of attributes, using their relative preferences for the underlying qualities without having to reveal them. The aggregate preferences, however, can be recovered in the form of average marginal causal effects (AMCEs) by regressing respondents' decision on each profile against the four attributes' values.

In study 2, which was conducted in rural counties of the Beijing province ($n = 892$), we modified the conjoint experiment in study 1 to include two new attributes indicating whether each township party secretary is known to be competent (Competence) or to have high moral values (Moral). Instead of rerandomizing the profiles, we presented each respondent with the same profiles in all six rounds, but at each round we hid some of the attributes and shuffled the remaining ones, so that respondents think the profiles are different each round. As illustrated in table C2, we hid information about the Competence attribute in one round, the Moral attribute in another, and then both these attributes in yet another round. If respondents' preference for anticorruption punishment operates through moral character evaluations, they would have to make inferences, or guesses, about the officials' moral qualities when making the final choice in conjoint rounds that hide the Moral attribute but not in rounds that explicitly assign a value for this attribute. As a result, the AMCE of Punishment in the former rounds would identify the entire *total effect* of Punishment, whereas the AMCE in the latter rounds would identify only a *direct effect*, because Punishment cannot change respondents' perception of the officials' moral qualities when information about them is already provided. The difference between these two quantities is the *indirect effect* of Punishment going through Moral, which captures the extent to which the preference for top-down punishment operates through punishment's influence on moral evaluations. Similarly, we can identify the indirect effect through competence evaluations as well as the joint indirect effect through both types of evaluations together.

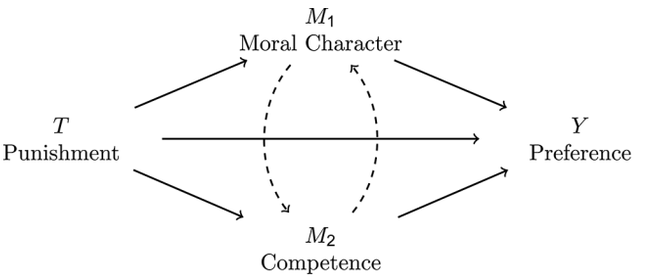

Figure 1. Causal mechanisms linking anticorruption punishment and citizens' preference.

Overall, our experimental design permits the identification and estimation of not only our respondents' preference for township officials who punish corruption by their subordinates but also the extent to which this preference arises out of beliefs about the officials' moral character or level of competence. We call this design, which builds on the parallel design for mediation analysis by Imai, Tingley, and Yamamoto (2013), the serial design. Given that our causal mechanisms are not necessarily independent, as illustrated by dashed arrows in figure 1, our design is better suited than other applications of the parallel design (e.g., Acharya, Blackwell, and Sen 2018). Specifically, for either the Moral or the Competence attribute, we can identify both the *natural indirect effect* that may operate through the other attribute as well as the *controlled indirect effect* when we block the other attribute by explicitly presenting information on it. Although we do not prefer one type of indirect effect over the other, distinguishing between them is important if our respondents' inferences about township officials' moral character and level of competence are somehow correlated. Appendix C discusses in detail the theories behind our design and its implementation.

Situating our study in authoritarian China is advantageous for many reasons. First, China seems to have a substantial amount of public support to explain (Chen 2004; Tang 2005), a sizable portion of which seems genuine (Pan and Xu 2018). Over the last two decades, government strategies of development and careful liberalization have generated this support. If we find that recent anticorruption initiatives also significantly shape public support in China, then they might carry even more weight in countries that have been less successful with other strategies. Second, real-world developments in China make a study of anticorruption's impact on public opinion increasingly relevant. Specifically, as economic growth slows and grassroots democratic reforms stagnate, anticorruption campaigns have become the main focus of the Chinese government, with increasingly frequent and intense punishment being meted out. Given the public's support for such punishment (Cunningham,

4. Appendix B (apps. A–G are available online) describes in detail the implementation of our experiment.

Table 1. Estimates of Indirect Effects of Punishment through Moral or Competence on Respondents' Binary Evaluation of a Township Party Secretary: Study 2

		Moral		Competence	
	Combined IE Rounds 1 and 6	Natural IE Rounds 1 and 3/4	Controlled IE Rounds 3/4 and 6	Natural IE Rounds 1 and 3/4	Controlled IE Rounds 3/4 and 6
Estimated IE	.070	.047	.024	.046	.023
	(.013)	(.073)	(.221)	(.073)	(.244)
Estimated total effect	.265	.265	.219	.265	.217
	(.000)	(.000)	(.000)	(.000)	(.000)

Note. Dependent variable: preferred choice. Randomization inference *p*-values in parentheses. IE = indirect effect.

Saich, and Turiel 2020, 9–10), it is reasonable to suspect that anticorruption punishment may help to substitute for or diversify the existing criteria used by citizens to evaluate authorities.

EXPERIMENTAL RESULTS

In both study 1 and study 2, we find strong and stable AMCEs for Punishment, suggesting that respondents do value top-down punishment by township party secretaries. All else being equal, township officials with a record of punishing corrupt village officials are preferred 20.3% of the time in study 1 and 26.5% in study 2. Although these effects are predictably smaller than those of Growth (35.7% and 32.1%), they are practically identical to those of Election (19.4% and 26.6%) and not statistically different from those of Welfare (25.3% and 21.1%).

An array of heterogeneous treatment effect analyses also finds that the AMCEs for the Punishment attribute are remarkably stable across different demographic groups.[5] In addition, they are practically unchanged even when we compare responses to profiles with only a "positive" value for one attribute and those with only a "negative" value for the same attribute, for each of the remaining attributes. The latter finding is particularly important with regard to the Growth attribute (see figs. D2 and D3), because it suggests that a township official's performance in promoting economic growth has no influence on respondents' preference for anticorruption punishment.

The mediation analysis, shown in table 1, provides support for both the deontological and instrumental mechanisms. Specifically, we find both the natural indirect effect going through moral character evaluations and the natural indirect effect going through competence evaluations to be statistically significant at the .1 level.[6] The magnitudes of these two indirect effects are almost identical and constitute about 18% of the total effect of Punishment. The joint indirect effect of both moral character evaluations and competence evaluations comprises about 26% of the total effect and is statistically significant at the .05 level.[7]

To confirm the robustness of this result, we apply the analysis to the five-point numerical rating that respondents gave to each individual profile, and we find the findings to be largely the same. As seen in table E1, the point estimates for the indirect effect going through moral character evaluations and through competence evaluations are now different, suggesting that their near identity above is not an artifact of our design.

Finally, to probe these findings' generality, we conduct study 3, which replicates study 2's design on an entirely different population through an online survey ($n = 1,152$). As evident in table A2, this sample is much more biased toward educated and wealthy urban respondents. The key findings, presented in appendix E, however, remain the same.

DISCUSSION AND CONCLUSION

In general, our studies produce a consistent set of results. First, although they are conducted on very different populations within China, our conjoint experiments all find a positive and significant AMCE for the Punishment attribute. This suggests that the preference for punishment of corruption is relatively stable across demographic groups in China.

5. Appendix D provides further details and discusses the heterogeneous effects in Kang and Zhu (2021).

6. We present *p*-values from randomization inference, which are conservative in experiments with multiple treatment effects such as ours. Note that our bootstrap-based *p*-values are consistently smaller than .05.

7. Appendix G discusses the relationship between different types of indirect effects.

Another striking finding is that citizens value anticorruption punishment at least as much as they do free and fair village elections, suggesting that anticorruption punishment can have as much potential to shape public opinion as these existing institutions for citizen participation. Punishment of corruption, however, consistently carries less weight than performance in promoting economic growth, which may provide a cautionary note to authoritarian leaders seeking to "recession proof" their support from the public.

In probing the causal pathway by which anticorruption punishment influences public opinion, we find that citizens prefer township officials who punish not only because punishment increases evaluations of competence but also because punishment of corruption signals a commitment to moral values. This result is consistent with our theory that anticorruption punishment matters to citizens not just instrumentally but also deontologically.

This article contributes to the theoretical and empirical literature on public opinion and citizen satisfaction by providing evidence for the causal pathways through which anticorruption initiatives might generate public support. The findings may extend to democracies as well, but they are most theoretically relevant for scholars of nondemocratic and hybrid regimes. For policy makers, these findings suggest that anticorruption punishment can build public support even when citizens have no evidence that it has decreased actual corruption, which may account for the rising popularity of illiberal, strongmen leaders around the world.

ACKNOWLEDGMENTS

We would like to thank Yingying Wen for making the field survey component of this project possible. Many colleagues provided invaluable feedback, including Daniel Corstange, Danny Hidalgo, Patrick Leslie, Adam Przeworski, Yiqing Xu, and Teppei Yamamoto. We also thank our anonymous reviewers for their feedback. Previous versions of this manuscript were presented at the 2017 annual meeting of the American Political Science Association in San Francisco, the Innovations in Political Methodology and China Study conference in Taipei, as well as in seminars at the Massachusetts Institute of Technology and New York University.

REFERENCES

Acharya, Avidit, Matthew Blackwell, and Maya Sen. 2018. "Analyzing Causal Mechanisms in Survey Experiments." *Political Analysis* 26 (4): 357–78.

Bies, Robert J. 1987. "The Predicament of Injustice: The Management of Moral Outrage." *Research in Organizational Behavior* 9:289–319.

Chen, Jie. 2004. *Popular Political Support in Urban China*. Washington, DC: Woodrow Wilson Center.

Cunningham, Edward, Tony Saich, and Jesse Turiel. 2020. "Understanding CCP Resilience: Surveying Chinese Public Opinion through Time." Technical report, Ash Center for Democratic Governance and Innovation, Harvard University.

Dickson, Bruce. 2016. *The Dictator's Dilemma: the Chinese Communist Party's Strategy for Survival*. New York: Oxford University Press.

Ferraz, Claudio, and Frederico Finan. 2008. "Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes." *Quarterly Journal of Economics* 123 (2): 703–45.

Festinger, Leon. 1950. "Informal Social Communication." *Psychological Review* 57 (5): 271–82.

Hainmueller, Jens, Daniel J. Hopkins, and Teppei Yamamoto. 2013. "Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments." *Political Analysis* 22 (1): 1–30.

Holmes, Leslie. 1993. *The End of Communist Power: Anti-corruption Campaigns and Legitimation Crisis*. Europe and the International Order. New York: Oxford University Press.

Imai, Kosuke, Dustin Tingley, and Teppei Yamamoto. 2013. "Experimental Designs for Identifying Causal Mechanisms." *Journal of the Royal Statistical Society* A 176 (1): 5–51.

Jiang, Junyan, and Dali L. Yang. 2016. "Lying or Believing? Measuring Preference Falsification from a Political Purge in China." *Comparative Political Studies* 49 (5): 600–634.

Kang, Siqin, and Jiangnan Zhu. 2021. "Do People Trust the Government More? Unpacking the Distinct Impacts of Anticorruption Policies on Political Trust." *Political Research Quarterly* 74 (2): 434–49.

Levi, Margaret, and Laura Stoker. 2000. "Political Trust and Trustworthiness." *Annual Review of Political Science* 3 (1): 475–507.

Li, Lianjiang. 2016. "Reassessing Trust in the Central Government: Evidence from Five National Surveys." *China Quarterly* 225:100–121.

Manion, Melanie. 2006. "Democracy, Community, Trust: The Impact of Elections in Rural China." *Comparative Political Studies* 39 (3): 301–24.

Meyer, Alexander, and Leah R. Rosenzweig. 2016. "Conjoint Analysis Tools for Developing Country Contexts." *Political Methodologist* 23 (2).

Pan, Jennifer, and Yiqing Xu. 2018. "China's Ideological Spectrum." *Journal of Politics* 80 (1): 254–73.

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

Wang, Erik. 2019. "Frightened Mandarins: The Adverse Effects of Fighting Corruption on Local Bureaucracy." SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3314508.

Wang, Yuhua, and Bruce Dickson. 2019. "How Corruption Investigations Undermine Regime Support: Evidence from China." SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3086286.

Whyte, Martin. 2010. *Myth of the Social Volcano: Perceptions of Inequality and Distributive Injustice in Contemporary China*. Stanford, CA: Stanford University Press.

Xi, Tianyang, Yang Yao, and Qian Zhang. 2018. "Dilemmas of Autocratic Governance: Theory and Evidence from the Anti-corruption Campaign in China." Working paper.