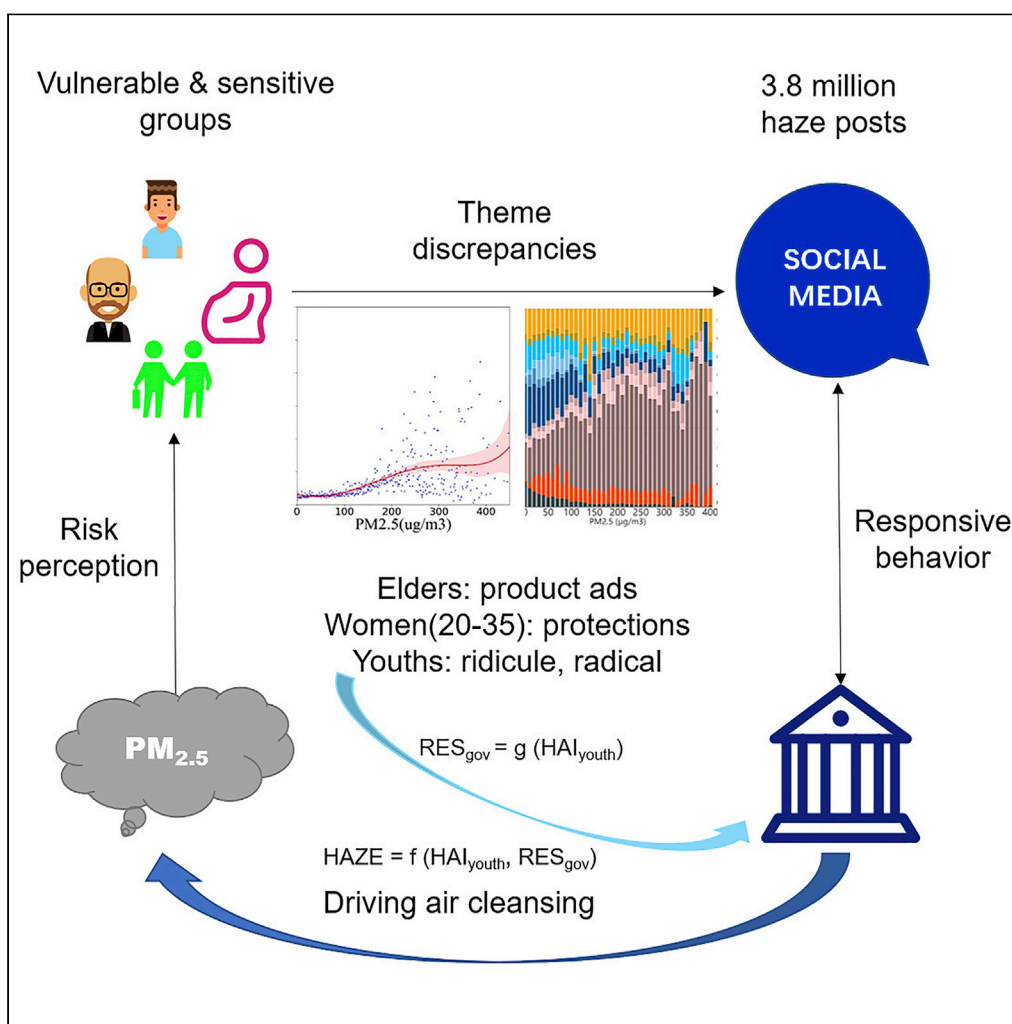


Article

Environmentally vulnerable or sensitive groups exhibiting varying concerns toward air pollution can drive government response to improve air quality



Z.H. Wang, W.H. Zhao, B. Wang, ..., Y.F. Sun, H. Shi, D.B. Guan

51022080@qq.com (B.W.)
zhangbin8706@163.com (B.Z.)
guandabo@hotmail.com (D.B.G.)

Highlights

Online haze concerns from environmentally vulnerable groups can drive air clean

Government response to their concern plays the key role in improving air quality

When $PM_{2.5}$ exceeds $200 \mu g/m^3$, their concern shifts from self-protection to governance

Elders are easily drawn by haze-related ads yet lack aware of scientific protection

Wang et al., iScience 25, 104460
June 17, 2022 © 2022 The Author(s).
<https://doi.org/10.1016/j.isci.2022.104460>



Article

Environmentally vulnerable or sensitive groups exhibiting varying concerns toward air pollution can drive government response to improve air quality

Z.H. Wang,^{1,2} W.H. Zhao,^{1,2} B. Wang,^{1,2,*} J. Liu,^{1,2} S.L. Xu,^{1,2} B. Zhang,^{1,2,*} Y.F. Sun,^{1,2} H. Shi,^{1,2} and D.B. Guan^{3,4,5,*}

SUMMARY

Air pollution seriously threatens human health, and its consequences are particularly prevalent among environmentally vulnerable or sensitive groups. However, whether the concerns among these groups are different and how they affect air pollution governance remain unclear. Here, we extract 3.8 million haze-related posts from China's Sina Weibo and analyze the concerns raised by these groups by constructing an air pollution notability index. The results show that protection is the key theme for women aged 20–35 years, while elderly individuals are easily influenced by haze-related product ads yet lack awareness of scientific-based protection. Concerns shared by young individuals are more effective in pressuring the government in cities that experience higher levels of pollution. Concerns shared by women are more effective in cities that experience lower levels of pollution. This study evidences the influence of the public concerns conveyed via social media on air pollution governance in China.

INTRODUCTION

In recent years, air pollution has become a major threat to people's quality of life, especially among environmentally vulnerable or sensitive groups in developing countries (WHO, 2019, 2019; Cerón-Bretón et al., 2018; Chowdhury et al., 2018; De Marco et al., 2019; Friedrich, 2018; Xue et al., 2019). Air pollution is seen as the greatest environmental threat according to the list of the top 10 threats to global human health released by the World Health Organization (WHO, 2019) in 2019, and according to the WHO, "Nine out of ten people in the world breathe polluted air every day" (WHO, 2019, 2019). Air pollution events have become a topic of public interest, as illustrated by the increasing number of related discussions across social media. Currently, people can easily and quickly express their feelings about and attitudes toward poor air quality because of the emerging development of social media platforms, such as Twitter (Mo and Hart, 2015; Wang et al., 2016), WeChat, Facebook, and Sina Weibo. These social media platforms allow people to provide real-time feedback about their concerns about environmental issues, such as the impact of Hurricane Harvey in 2017 in Texas (Liu et al., 2018; Yuan and Liu, 2018) and the wildfires in 2018 in California (Cisneros and Schweizer, 2018). Sina Weibo is the largest post platform in China, with more than 462 million users, accounting for 55.7% of the total number of internet users. In 2018, every day, more than 200 million active users on average shared their views on Sina Weibo (Weibo, 2019).

Ongoing research has taken an increased interest in public concern about air pollution (Chuanshen et al., 2018; North et al., 2014; Schusky, 1966; Tvinnereim et al., 2017; Xu et al., 2018; Zhang and Xiaowen, 2017). In the face of air pollution issues, women aged between 20 and 35 years, as well as young and elderly individuals, are seemingly more vulnerable or sensitive than are other groups, and their perceptions of air pollution risks tend to be higher than those of others (Hunter et al., 2004). The concerns of these individuals are therefore of particular interest.

Scholarly interest in vulnerable groups originated with social exclusion theory. Individuals' participation in public debates, to articulate their own interest, is resource-demanding (Fahmy et al., 2018; Marien et al., 2010), with vulnerable groups at a disadvantage given their limited social, political, legal, and other types of resources; they often lack the opportunity for political participation and cannot easily access suitable channels through which to express their interests. In the process of formal political participation, these

¹School of Management and Economics, Beijing Institute of Technology, Beijing 100081, China

²Research Center for Sustainable Development and Intelligent Decision, Beijing Institute of Technology, Beijing, China

³Department of Earth System Science, Tsinghua University, Beijing 100083, China

⁴The Bartlett School of Sustainable Construction, University College London, London, UK

⁵Lead contact

*Correspondence: 51022080@qq.com (B.W.), zhangbin8706@163.com (B.Z.), guandabo@hotmail.com (D.B.G.)

<https://doi.org/10.1016/j.isci.2022.104460>



groups are often “ignored and excluded by society” and thus become environmentally vulnerable (Jiang et al., 2017).

In practice, informal political participation through the “media” has become an important channel through which vulnerable groups can express their rights and interests (Fahmy et al., 2018; Mo and Hart, 2015). Currently, online social media provides relatively convenient and low-cost channels through which to help vulnerable groups express their own interests (Liu and Zhao, 2017). With governance modernization and internet development, Chinese citizens have become accustomed to expressing their demands via media in daily life, with the government being responsive to the will of the people (Distelhorst and Hou, 2014; Hansen and Liu, 2018; Meng et al., 2017). However, the extent to which the concerns of vulnerable or sensitive groups can play a role in protecting their environmental rights and interests remains underexplored.

A substantial body of work has shown that election pressure is the main driver of government responsiveness (Tang et al., 2020). However, a growing number of empirical studies focusing on non-Western electoral systems argue that the government is responsive to the will of people regardless of election pressure (Distelhorst and Hou, 2014; Hansen and Liu, 2018). The Chinese government has been actively creating various types of e-participation initiatives, including online petition forums (local leader message boards) and government-sponsored social media accounts, to communicate air pollution issues to the public (Meng et al., 2015). In response to public concerns, the government has taken further measures to control air pollution by increasing its investment in air pollution control and promulgating relevant laws and regulations. For example, to reduce the air pollution caused by automobile exhaust and industrial waste gas emissions, which were concerns raised by the public, vehicle-restricted policies and restrictions on pollutant discharge have also been implemented by the government.

Most previous studies on public concerns are based on survey (Fan et al., 2018; Meng et al., 2015; Pantavou et al., 2018; Siamak et al., 2018) or statistical data (Howel et al., 2010; Lu et al., 2018; Shi and Chen, 2019; Tran et al., 2018), and although some scholars have carried out research on public concerns regarding environmental issues using social media data, only a few studies have focused on the content of microblogging posts that reflect the concerns of environmentally vulnerable or sensitive groups (Zheng et al., 2019; Wang et al., 2020). To date, there remains a lack of understanding of the concerns from these groups, and there is insufficient knowledge regarding whether these concerns are different than those of other groups, and how such different concerns affect air pollution governance in China.

Here, based on a conceptual framework established through the theory of risk perception and the theory of planned behavior (TPB), the disparities in terms of public concerns regarding air pollution across different environmentally vulnerable or sensitive groups in social media have been explored, and the interaction mechanisms among the voices of these groups, government responses, and air quality improvement have been investigated. A crowd-based air pollution notability index is constructed to quantitatively measure public concern using the latent Dirichlet allocation (LDA) model. The results show that when $PM_{2.5}$ concentrations exceed $200 \mu\text{g}/\text{m}^3$ (the maximum 24-h average $PM_{2.5}$ exposure level should not exceed $75 \mu\text{g}/\text{m}^3$ according to China’s National Ambient Air Quality Standard), people’s concerns shift from self-protection to air pollution accountability. Particularly, concerns from these groups have been found to have had a differential influence on air pollution governance in China through the mediation role played by government response. This paper adds to the literature on the relationship between environmentally vulnerable or sensitive groups and air quality improvement, which can help promote the modernization of air pollution governance in the internet era through a bottom-up approach (Cheng et al., 2017).

RESULTS

Public concerns vary according to air quality levels

Figure 1 shows the correlation between public air pollution notability and $PM_{2.5}$ concentration. The average air pollution notability is 382. Overall, the worse the air quality is, the higher the corresponding air pollution notability. Public air pollution notability reaches 566 on average during the days with $PM_{2.5}$ concentrations over $200 \mu\text{g}/\text{m}^3$ (grade 5, heavily polluted), which is more than two times higher than that on the days with $PM_{2.5}$ concentrations below $75 \mu\text{g}/\text{m}^3$ (grade 2, good air quality). Air pollution alarm, where levels of $PM_{2.5}$ reach hundreds and exceed the alarm levels, plays a significant role in raising public air pollution notability during this period. When the $PM_{2.5}$ concentration is approximately $250 \mu\text{g}/\text{m}^3$, public air pollution notability jumps to over 1700, which is approximately four times higher than the overall average level.

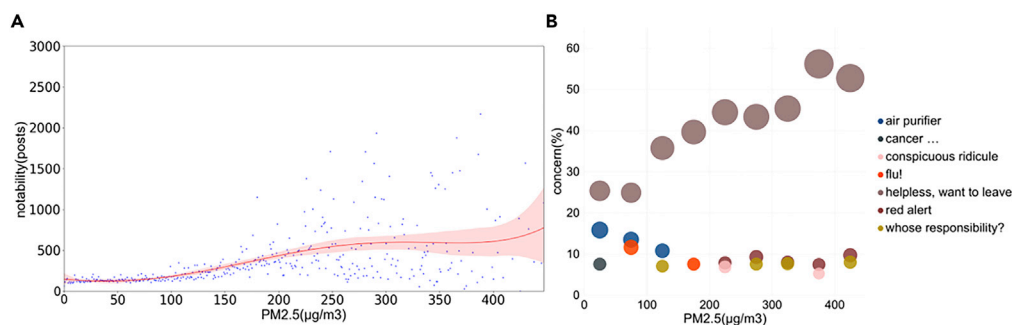


Figure 1. The trend of air pollution notability and top air pollution theme of public concern in response to PM_{2.5} concentration

Figure 1A shows the trend of air pollution notability in response to PM_{2.5} concentrations. The vertical axis is air pollution notability. Figure 1B shows the air pollution of concern across different levels of PM_{2.5} concentration. The vertical axis is the air pollution of public concern. The measurement methodology for both air pollution notability and air pollution of concern is shown in Equations 1, 2, and 3 in the *Experimental Procedures* section. The line is the total air pollution notability under the corresponding PM_{2.5} concentration. The bubbles refer to the public concern for different types of air pollutions, which are identified from the sample of data from Sina Weibo from January 2017 to December 2018 using the LDA model shown in the *Experimental Procedures* section. The top three air pollutions in each interval of PM_{2.5} concentration are shown in Figure 1B.

In terms of public concern, “Helpless, want to leave (无奈, 想走)” is the hottest topic (see Table S4 for theme definition) across the haze-related posts on *Sina Weibo*, no matter what the state of the air quality is. The average public concern expressed as “Helpless, want to leave” is 39%, which is nearly six times higher than the average public concern related to “Red alert (红色预警)”, the second hottest topic.

“Helpless, want to leave” reflects people’s frustration at the expressed desire but failure to escape from areas where there is a serious air pollution issue, while facing the haze weather. There is also an increasing trend of “Helpless, want to leave” alongside the growth of PM_{2.5} concentration. Public concern related to “Helpless, want to leave” is around 20% when PM_{2.5} concentration is below 50 µg/m³. However, when the PM_{2.5} concentration raises from 110 µg/m³ to 360 µg/m³, the public concern in terms of “Helpless, want to leave” rose from 30% to 50%. People’s feelings about “Helpless, want to leave” become more intense with the increase in the intensity of air pollution.

The public concern expressed as “Air purifier (空气净化器)” is around 15% when the PM_{2.5} concentration is below 150 µg/m³, which is second to the “Helpless, want to leave” hot topic during haze weather condition. When the PM_{2.5} concentration exceeds 200 µg/m³, the public concern “Air purifier” drops to 5%. Comparatively, when the PM_{2.5} concentration is below 150 µg/m³, the public concern in terms of “Whose responsibility? (治理不力, 谁负责?)” is 6%, but rises to 10% when PM_{2.5} concentration exceeds 260 µg/m³. These findings show that Chinese citizens prefer to choose their own independent protection measures rather than rely on the government or society, when the level of air pollution is tolerable. However, when the PM_{2.5} concentration exceeds the alarm levels (>200 µg/m³), self-protection can no longer solve the problem. People have to find other ways to address the air quality problems, and question who should take responsibility for the air pollution: the government or large-scale emission enterprises?

Additionally, public concern expressed as “conspicuous ridicule (炫耀性吐槽)” exceeds 7% when PM_{2.5} concentration is between 150 µg/m³ and 200 µg/m³, which is second to “Helpless, want to leave”. However, this figure decreases to below 4% during other air quality conditions, meaning that it drops out of the TOP three hottest themes, as shown in Figure 1. “Conspicuous ridicule” reflects that people from areas with good air quality show off to the people who are suffering from poor air quality. The related posts on *Sina Weibo* show how those experiencing good air quality in places like Chengdu and Hainan complain about the air pollution in places like Beijing. According to the conspicuous consumption theory (Bernheim, 1996; Woodside, 2012), showing off behavior often happens in two comparable groups without too much disparity. If the regional air quality gap is high, people from good air quality areas are less concerned about showing off and are more likely to display other emotions, such as sympathy. Some posts on *Sina Weibo* convey sentiment such as “hope people can get better protection in Beijing”,

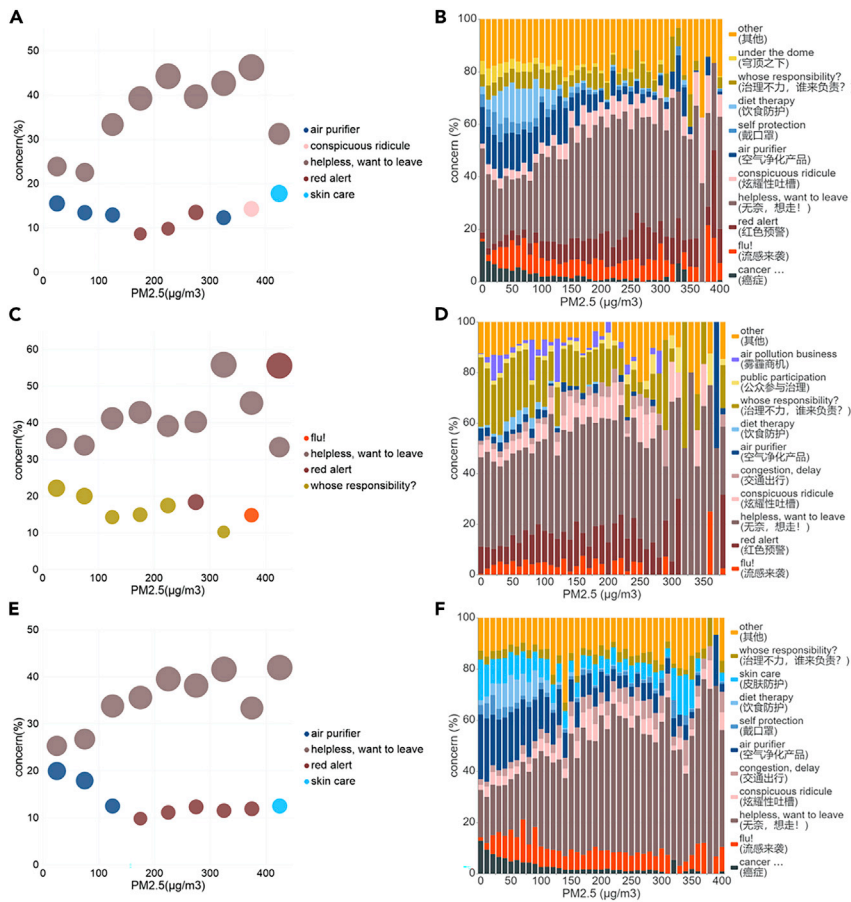


Figure 2. Themes of concern by subgroup

Figures 2A, 2C, and 2E show the top two types of air pollution in terms of expressed concern in response to PM_{2.5} concentration for the young individuals, elderly individuals, and women aged between 20 and 35 years, respectively. Figures 2B, 2D, and 2F show the ratios of public concern themes for the corresponding vulnerable or sensitive groups. The x axis represents the air pollution concentration, and the y axis represents the average of the group's theme notability in each specific air pollution concentration range; different bubble colors indicate different themes, and the size of the bubble is proportional to the theme notability in that specific PM_{2.5} concentration range.

“wish there would be better air quality for Beijing people”, and “worry about the children living in places with air pollution.”

Discrepancies in concerns on air pollution across groups

We further classify three groups of people vulnerable or sensitive to air pollution (namely the youths, the elders, and women aged between 20 and 35) (<http://data.stats.gov.cn/easyquery.htm?cn=C01>). Detailed definitions of each group are shown in the Experimental Procedures section. From that, we investigate their air pollution notability and themes of their concerns toward the variety of PM_{2.5} concentrations.

For the women aged between 20 and 35, as shown in Figure 2, protection is the key word. Topics like “Air purifier(空气净化器)”, “Self-protection!(自我保护)”, “Diet therapy!(饮食防护)”, and “Skin care(皮肤防护)” are all protection-related themes and account for 24.7% of air pollution theme concern in the female aged between 20 and 35 group. This figure is 1.5 times higher than the average level of the total group. It indicates that women aged between 20 and 35 are concerned more about self-protection than the average in haze weather. Moreover, this group pays more attention to self-protection in less polluted weather, in comparison to other groups. When PM_{2.5} concentration is below 200 µg/m³, the concern on protection-related themes in the female aged between 20 and 35 group is 30% on average. This figure is 1.5 times higher than the average level of the total group. Additionally, “Helpless, want to leave” is the hottest theme

for females aged between 20 and 35, with related concern of 41% on average. This figure is about 1.1 times higher than the average level of the total group, showing that they have a greater tendency to escape from the city experiencing air pollution to better protect themselves. In particular, in high-level air pollution scenarios where $PM_{2.5}$ concentration is higher than $250 \mu\text{g}/\text{m}^3$, the concern of females aged between 20 and 35 in terms of “Helpless, want to leave” increases to 50% on average. Comparatively, their concern on protection-related themes decreases to 8%. A considerable proportion of protective topics shifted to “Helpless, want to leave” in the scenario of severe pollution weather. It indicates that females aged between 20 and 35 feels unsafe when the area they are in suffers from severe air pollution (Jasemzadeh et al., 2016), and shows a tendency to flee the city.

Compared to other groups, the youths (age < 18) pay more attention to the theme of “conspicuous ridicule”. Their concern in terms of “conspicuous ridicule” is 13% on average when $PM_{2.5}$ concentration exceeds $350 \mu\text{g}/\text{m}^3$, which is 1.5 times higher than the average level of the total group. This is particularly noticeable when $PM_{2.5}$ concentration exceeds $360 \mu\text{g}/\text{m}^3$, with the concern on “conspicuous ridicule” reaching 22% for the youths. Showing-off behavior has an important role for young people in endeavoring to get attention from others (Koyama and Smith, 1991), resulting in a tendency to compare areas where air pollution is good or bad. Additionally, “Helpless, want to leave” is also the hottest theme in the adolescent group, but the trend is smoother than in other groups. The average growth rate of public concern of the youth group in terms of “Helpless, want to leave” is 2.5%, which is three-quarters of the average level of the total group. Although there is also a strong desire for the youths to leave the city experiencing air pollution, it cannot be as easily realized as by other groups due to their limitations in being able to afford to live in another city. Moreover, youths pay much more attention to the theme “Public opinion events (舆情事件)”. Their concern related to “Public opinion events” is nearly 1.7%, which is almost three times higher than the average level of the total group. This indicates that during air pollution conditions, the youth group prefers to discuss haze-related news, and quote past air pollution events.

The elders can be described as a group “believing in advertising but neglecting science”. Compared to other groups, the concern related to “air pollution business (雾霾商机)” is 3%, which is three times higher than the average level of the total group and is in the TOP 10 hottest themes. It indicates that the elders are more likely to be attracted by haze-related advertisements and products. Comparatively, the elders pay less attention to self-protection. Their focus on “air purifiers” and “Diet therapy” was 3% and 0.9%, respectively, which are 50% lower than the average level of the total group. These results indicate that awareness of self-protection against air pollution is relatively weak in the elders. In China, it is common to see a large number of aged people doing outdoor activities such as outdoor dancing and Tai Chi in conditions of smog (Liang, 2014). Moreover, the elders pay more attention to the theme of “Whose responsibility?” than other groups. The concern of the elders related to this topic is 15.1%, while the overall group level is only 6.7%. These patterns show that the elders are more concerned about the environmental governance of air pollution and are anxious about the reasons behind, and responsibility for, the poor governance of air pollution. When they are suffering from bad air quality, they rely more on external help rather than self-protection. In addition, at high levels of air pollution ($PM_{2.5}$ concentration $>250 \mu\text{g}/\text{m}^3$), concentration of public concern in terms of “Helpless, want to leave” among the elders is 50% on average, which is 5%–10% higher than the average level of the total group. It shows that the elders are more likely to choose to stay away from cities experiencing air pollution when haze is severe. More and more elders are buying real estate in southern cities such as Sanya, Shenzhen, and other cities in China that have good air quality (W.J., 2017).

Role of public notability in improving air quality

To identify whether public air pollution notability leads to reduced local air pollution, we build a regression model by matching the monthly air quality data with the public notability data generated from Sina Weibo for 322 cities in China. We focus, in particular, on the effects of air pollution notability in different environmentally vulnerable or sensitive groups.

The total influence coefficient of the public notability level on the $PM_{2.5}$ concentration is -0.062 , which is significant at the 1% level (coefficient = -0.062 , p value < 0.01). This result means that with a 1% increase in haze-related posts in each city per year, the average annual $PM_{2.5}$ concentration decreases by 0.062% (see Table 1 for details). Aiming to eliminate the influence of heating differences on air quality in China, we conduct a subgroup analysis to ensure the robustness of the results by dividing the samples into southern and northern China, which are shown in the Table S5 Robustness check 1. The results are robust in both regions, regardless of whether or not there is central heating.

Table 1. Regression results for the relationship between air pollution notability and air quality improvement

	Model ALL	Model G1	Model G2	Model G3
Group effect	ln(HAZE_{t+n,d})			
ln(HAI _{t,d})	−0.06241*** (0.00600)	−0.05855*** (0.00834)	−0.06218*** (0.00892)	−0.10900*** (0.02673)
ln(Income _{t,d})	−0.07523*** (0.01153)	−0.05644*** (0.01480)	−0.08078*** (0.01938)	−0.25994*** (0.05356)
ln(Industry _{t,d})	0.44975 (0.81008)	1.24846*** (0.26659)	0.00000 (.)	−1.77096*** (0.28035)
ln(Wind _{t+n,d})	0.02336** (0.01040)	0.02327 (0.01428)	0.01976* (0.01053)	0.03805 (0.18770)
ln(Rain _{t+n,d})	−0.05771*** (0.00680)	−0.07071*** (0.01314)	−0.04987*** (0.00902)	−0.04986*** (0.01722)
_cons	3.07029 (3.04345)	−0.08740 (0.98368)	4.85068*** (0.16535)	13.18746*** (1.27709)
Season	YES	YES	YES	YES
City	YES	YES	YES	YES
Adj_R2	0.7379	0.6942	0.7267	0.7306
N	2040	840	1020	180

Notes: *p < 0.10, **p < 0.05, and ***p < 0.01. Standard errors are in parentheses. Standard errors are clustered at the city quarterly level. See Tables S5 and S6 for Robustness check.

Additionally, the marginal effect of the public notability level on air quality improvement is more significant in cities with more air pollution exposure. The results of the subgroup regression considering the air pollution level of each city are displayed in Table 1. The marginal effect in cities that experience the longest (more than 6 months) air pollution exposure (coefficient = −0.109, p value < 0.01) is 1.85 times that of the cities that experience the lowest (less than 4 months) air pollution exposure (coefficient = −0.059, p value < 0.01). The results are robust in both northern and southern China cities that experience different levels of air pollution exposure (see Table S2 Robustness check two for details), regardless of whether or not there is central heating.

Considering the concerns of different environmentally vulnerable or sensitive groups, the effects of their notability on improving air quality vary (see the Table 2 for details). The public notability of the young individuals exhibits an effect on air quality improvement in cities that experience longer air pollution exposure (more than 6 months, coefficient = −0.071, p value < 0.05). Comparatively, in cities that experience medium (between 4 and 6 months) and low (less than 4 months) air pollution exposure, the significant effect (coefficient = −0.071, p value < 0.01 and coefficient = −0.057, p value < 0.01, respectively) comes from the public notability of women aged between 20 and 35 years.

A possible reason for this is that women aged between 20 and 35 years are more concerned about air pollution even if the air quality is relatively acceptable. Their number of haze-related posts is 1.9 times and 24.6 times higher than that of young and elderly individuals, respectively. Comparatively, the public concern of the young individuals is only high when air pollution is serious (more than 6 months exposure to air pollution). Particularly, their expressions and views are more emotional and critical; they preferentially process negative information (Gao and Jie, 2015), easily express radical emotions such as anger, and can become dispirited when suffering bad or uncomfortable experiences. To provide statistical proof that Sina Weibo posts by young individuals are more emotional than those by elderly individuals and women aged between 20 and 35 years, we randomly select posts from these three groups. The emotional level of each post is evaluated by a five-point measurement scale, the results of which show that the emotional score of young individuals is 10 and 21 times higher than that of elderly individuals and women aged between 20 and 35 years (see Figure S2 for details). As governments often pay attention to emotional comments, which can affect social stability, the emotions of young individuals on social media can ultimately explain improvements in air quality. This aspect is especially true in China, given that social stability and harmonious

Table 2. Effect of notability of different groups on air quality improvement

Group effect	Model G1M1 $\ln(\text{HAZE}_{t+n,d})$	Model G2M1	Model G3M1
$\ln(\text{HAI}_{t,d_young})$	−0.00115 (0.01310)	0.00345 (0.01099)	−0.07092** (0.03159)
$\ln(\text{HAI}_{t,d_aged})$	0.03056 (0.02119)	0.02726 (0.01772)	0.00046 (0.02944)
$\ln(\text{HAI}_{t,d_exmother})$	−0.05663*** (0.01422)	−0.07058*** (0.01167)	−0.04161 (0.04139)
$\ln(\text{Income}_{t,d})$	−0.04049*** (0.01457)	−0.07213*** (0.01947)	−0.26777*** (0.05061)
$\ln(\text{Industry}_{t,d})$	1.45394*** (0.26555)	0.00000 (.)	−1.78192*** (0.38182)
$\ln(\text{Wind}_{t+n,d})$	0.01476 (0.01327)	0.01682 (0.01065)	0.01529 (0.17209)
$\ln(\text{Rain}_{t+n,d})$	−0.07469*** (0.01331)	−0.05009*** (0.00896)	−0.05214*** (0.01696)
_cons	−1.05266 (0.97725)	4.68306*** (0.15497)	13.06375*** (1.63037)
Season	YES	YES	YES
City	YES	YES	YES
Adj_R2	0.6900	0.7278	0.7335
N	840	1020	180

Notes: *p < 0.10, **p < 0.05, and ***p < 0.01. Standard errors in parentheses. Standard errors are clustered at the city quarterly level.

development are state priorities (Leeson, 2006). The effect of public concern on air quality in cities that experience low- and medium-level air pollution exposure is largely due to the number of posts published by women aged between 20 and 35 years. When air pollution becomes serious, emotional posts by young individuals are more attractive for air cleansing and cover the effects of other groups.

Mechanism of public concern, government response, and cleaner air

Government response is one of the missing links in the relationship between public concern and air quality improvement. According to the theory of government responsiveness, the government is sensitive to people's environmental demands (Li et al., 2019), which may lead to air quality improvement (Figure 3). Here, 328,192 haze-related government responsive posts are extracted from Sina Weibo (Figure S3). We use mediation effect analysis to explore the mechanism by which public notability improves air quality from the perspective of government response (see Figure S4 for details). The results of Model M1 show that public notability has a significant effect on air quality. On this basis, we further analyze whether the effect works through the government response to public concern.

As shown in the Figure S4, there is a positive correlation between public notability level and government response, which is significant at the 1% level (coefficient = 0.079, p value < 0.01). This finding indicates that with a 1% increase in the number of haze-related posts per year in each city, the average number of government responses on Sina Weibo increases by 0.079%. Through mediating effect analysis (see Table 3), government response has a negative effect on air pollution (coefficient = −0.023, p value < 0.01). Furthermore, we add the government response variable into Model M3 and find that the influence coefficient (coefficient = −0.060, p value < 0.01) of public notability on air quality decreases after including the mediating variable. Moreover, the model indicates that government response plays a partial mediating role in the above relationship. Public notability affects air quality due to the mediating role played by the government.

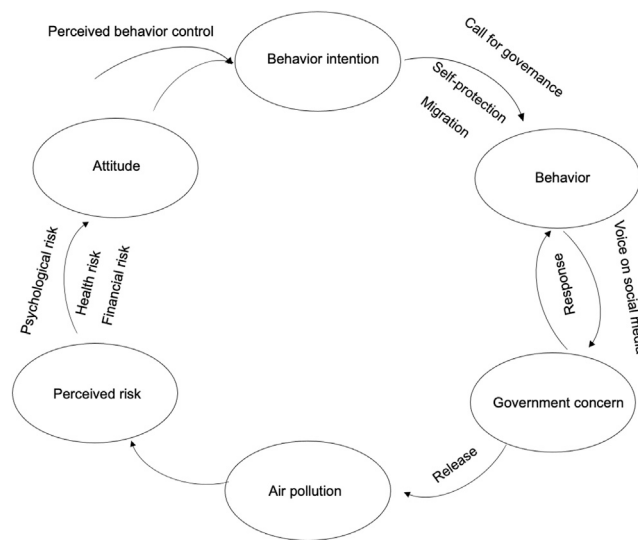


Figure 3. Theoretical framework

These results demonstrate that the online interaction between the government and public voices reflecting air-pollution-related problems can be heard and addressed. Higher air pollution notability attracts more government attention, thus pushing the government toward finding more ways to deal with air pollution problems (Reed et al., 2014). The government responds to public notability via online channels and takes further actual measures to control air pollution (Che et al., 2011; Li et al., 2019; Zhang et al., 2020). In terms of air pollution management, local governments have introduced air pollution control measures (Wang and Watanabe, 2019; Zhang et al., 2018). For example, to reduce the air pollution caused by automobile exhaust, vehicle-restricted policies have been introduced. Overall, these expressions of public concern are possible drivers of government attention, given the need to secure social stability, with potential effects on air quality improvements (Leeson, 2006).

DISCUSSION

This work extracts 3.8 million air pollution posts from a total of 3.5 billion Sina Weibo posts from January 2017 to September 2018. Based on the collected big data, we identify air pollution themes on this social media platform and explore the relationship between air pollution notability and air quality improvement, particularly among environmentally vulnerable or sensitive groups such as women aged between 20 and 35 years, as well as young and elderly individuals. A conceptual model is created based on the theory of risk perception, the TPB and government responsiveness to assess the mechanism of public concerns, government response, and air quality improvement.

The results show that high levels of air pollution notability on Sina Weibo are related to poor air quality conditions. Opinions about whether to stay or leave a city when suffering from air pollution that reaches the level of $110 \mu\text{g}/\text{m}^3$ is mostly advocated via Weibo posts, with expressions of helplessness and frustration coming from those facing air pollution. People are more concerned about self-protection when air pollution is bearable (when the $\text{PM}_{2.5}$ concentration is below $150 \mu\text{g}/\text{m}^3$) but rely on air pollution governance if air pollution becomes serious (when the $\text{PM}_{2.5}$ concentration exceeds $200 \mu\text{g}/\text{m}^3$). The air pollution themes among the public opinions posted on Sina Weibo vary across different groups. Women aged between 20 and 35 years care more about self-protection from air pollution, while young individuals show more sensitivity to past air pollution events and prefer to show off the better conditions in which they live when they are in places with good air quality. Elderly people are easily influenced by air pollution advertisements and have a lower awareness of scientific-based protection from air pollution.

Public notability plays a significant role in improving air quality through the mediating role of the government. Particularly in cities that experience air pollution for more than 6 months, public concern from the young group helps strengthen the effects of improving air quality. The expressions among

Table 3. Mechanism analysis for the relationship between the air pollution notability level and air quality improvement

	Model M1	Model M2	Model M3
Mediation effect	$\ln(\text{HAZE}_{t,d})$	$\ln(\text{RES}_{t,d_government})$	$\ln(\text{HAZE}_{t,d})$
$\ln(\text{RES}_{t,d_government})$			-0.02335** (0.01042)
$\ln(\text{HA}_{i,t,d})$	-0.06241*** (0.00600)	0.07906*** (0.01299)	-0.06049*** (0.00596)
$\ln(\text{Income}_{t,d})$	-0.07523*** (0.01153)	0.04541 (0.03032)	-0.07416*** (0.01134)
$\ln(\text{Industry}_{t,d})$	0.44975 (0.81008)	2.01043 (1.80993)	0.49344 (0.84474)
$\ln(\text{Wind}_{t+n,d})$	0.02336** (0.01040)	-0.00704 (0.02188)	0.02267** (0.01060)
$\ln(\text{Rain}_{t+n,d})$	-0.05771*** (0.00680)	-0.06596*** (0.01056)	-0.05735*** (0.00676)
_cons	3.07029 (3.04345)	-4.86336 (6.79746)	2.96236 (3.17506)
Season	YES	YES	YES
City	YES	YES	YES
Adj_R2	0.7379	0.8475	0.7384
N	2040	2040	2040

Notes: *p < 0.10, **p < 0.05, and ***p < 0.01. Standard errors in parentheses. Standard errors are clustered at the city quarterly level. See Figure S4 for Result of the mediation effect based on causal step regression. The results from Models M1, M2, and M3 all show that government response plays a channel role in the relationship between air pollution notability and air quality improvement.

this group posted on social media are more emotional or radical than are those of the other population groups. It is hopeful to see that young people, as bearers of the times, have begun to take more sense of social responsibility. The concerns among those in the youth group should be treated positively and rationally.

With the development of the internet, social media has gradually turned into a channel carrying all kinds of information and functions in addition to social features. When the influence of social media is valued in the environmental governance domain, it is not only a propaganda channel through which to disseminate political information but also a platform on which citizens can participate in governance. This study shows that the voices of vulnerable or sensitive groups can be heard, which then raises governmental concern and eventually results in improved air quality. These findings lend support to the view that internet and information technologies can be a tool with which to empower citizens' opinions and rights in the digital age. The public, especially vulnerable groups, should be encouraged to use social media channels to actively communicate with and influence the government with the view of supporting the modernization of national governance through a bottom-up approach. Specifically, considering that air pollution notability is apparently different under different air pollution levels, it is important to choose the proper timing of facilitating citizens' engagement in air pollution governance. For example, when the PM_{2.5} concentration exceeds 200 μg/m³, with the obvious increase in public air pollution notability, policymakers may collect more sufficient and efficient information on air pollution governance. In addition, as we have found in this paper, under different haze concentrations, the most concerned themes or demands of different groups are different. Thus, after considering the demands of the vast majority of people via social media, the demands of vulnerable groups may require special attention. In future research, it will be necessary to conduct more detailed research on governance mechanisms and policymaking strategies based on more detailed demography and social media data. Otherwise, the environmental rights and interests of the vulnerable groups may not be considered fairly.

Limitations of the study

Although Sina Weibo covers approximately 55% of netizens during the research period and despite mounting evidence of the advantages of social media data representing public opinion, this study faces data limitations that need to be overcome in future research. For example, the group of women aged between 20 and 35 years is selected as representatives of a class of environmentally sensitive groups, with the majority of pregnant women falling into this group. This group cannot represent pregnant women but can reflect the characteristics of those in the pregnancy-related group to some extent. The other issue is the time span of the data. The algorithm to calculate group notability and group-related themes of concern relies on daily posts at the group-city level; however, we have no such data before 2017, which limits the time span of the current study. Longer term mechanisms and the effects of public concern on air quality improvement can be explored in future research.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
 - Lead contact
 - Material availability
 - Data and code availability
- **METHOD DETAILS**
 - Theoretical framework
 - The air pollution theme identification based on the LDA model
 - Experiments for identifying the emotional level of microblogs in different populations
 - Definition and measurement of air pollution notability
 - Effect of public concern on air quality improvement
 - Robustness check
 - Mechanism analysis
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - Data source of air pollution public concern and air pollution

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2022.104460>.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Reference No. 72141302, 72074026, 71625003 and 72140002), Major Research Projects in Philosophy and Social Sciences of the Ministry of Education (No. 21JZD027), Philosophy and Social Science Foundation of Beijing (No. 21GLC057), and Beijing Natural Science Foundation of China (No. 9212016).

AUTHOR CONTRIBUTIONS

Z.H.W., D.B.G., B.W., and B.Z. designed the study. W.H.Z., B.W., and J.L. completed the LDA-model-related work, constructed the indexes, and visualized the data. Y.F.S., B.W., B.Z., and H.S. completed the econometric-model-related work. Z.H.W., Y.F.S., W.H.Z., and S.L.X. wrote the first draft. Z.H.W., D.B.G., B.W., B.Z., W.H.Z., Y.F.S., S.L.X., and J.L. contributed to the interpretation of the results and revision of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

The author list of this paper includes contributors from the location where the research was conducted who participated in the data collection, design, analysis, and/or interpretation of the work.

Received: March 18, 2022

Revised: April 11, 2022

Accepted: May 16, 2022

Published: June 17, 2022

REFERENCES

- Bernheim, L. (1996). Veblen effects in a theory of conspicuous consumption. *Am. Econ. Rev.* 86, 349–373.
- Che, W., Zheng, J., Wang, S., Zhong, L., and Lau, A. (2011). Assessment of motor vehicle emission control policies using Model-3/CMAQ model for the Pearl River Delta region, China. *Atmos. Environ.* 45, 1740–1751.
- Cheng, Z., Li, L., and Liu, J. (2017). The emissions reduction effect and technical progress effect of environmental regulation policy tools. *J. Clean. Prod.* 149, 191–205.
- Cerón-Bretón, R., Cerón-Bretón, J., Lara-Severino, R.C., Espinosa-Fuentes, M.L., and Uc-Chi, M.P. (2018). Short-term effects of air pollution on health in the metropolitan area of Guadalajara using a time-series approach. *Aerosol Air Qual. Res.* 18.
- Chowdhury, S., Dey, S., and Smith, K.R. (2018). Ambient PM_{2.5} exposure and expected premature mortality to 2100 in India under climate change scenarios. *Nat. Commun.* 9, 318. <https://doi.org/10.1038/s41467-017-02755-y>.
- Chuanshen, Q., Xu, J., Wong-Parodi, G., and Xue, L. (2018). Change in public concern and responsive behaviors toward air pollution under the dome. *Risk Anal.* 40, 1983–2001.
- Cisneros, R., and Schweizer, D.W. (2018). The efficacy of news releases, news reports, and public nuisance complaints for determining smoke impacts to air quality from wildland fire. *Air Qual. Atmos. Health* 11, 423–429.
- De Marco, A., Proietti, C., Anav, A., Ciancarella, L., D'Elia, I., Fares, S., Fornasier, M.F., Fusaro, L., Gualtieri, M., Manes, F., et al. (2019). Impacts of air pollution on human and ecosystem health, and implications for the National Emission Ceilings Directive: Insights from Italy-NC-ND. *Environ. Int.* 125, 320–333. <https://doi.org/10.1016/j.envint.2019.01.064>.
- Distelhorst, G.H., and Hou, Y. (2014). Ingroup Bias in official behavior: a national field experiment in China. *Q. J. Political Sci.* 203–230.
- North, M.L., Alexis, N.E., Ellis, A.K., and Carlsten, C. (2014). Air pollution and asthma: how can a public health concern inform the care of individual patients? *Ann. Allergy Asthma Immunol.* 113, 343–346.
- Fahmy, E., Sutton, E., and Pemberton, S. (2018). Understanding social exclusion: the views of the UK public. *J. Poverty Soc. Justice*, 439–458.
- Fan, Y., Ding, L., Liu, C., Xu, L., Stephen, N., and Wang, J. (2018). Haze attitudes and the willingness to pay for haze improvement: evidence from four cities in Shandong province, China. *Int. J. Environ. Res. Publ. Health* 15. <https://doi.org/10.3390/ijerph15102297>.
- Friedrich, M.J. (2018). Global impact of air pollution on children's health. *JAMA* 320, 2412. <https://doi.org/10.1001/jama.2018.19559>.
- Gao, and Jie. (2015). Political rationality vs. technical rationality in China's target-based performance measurement system: the case of social stability maintenance. *Policy Society* 34, 37–48.
- Hansen, M.H.L., and Liu, Z. (2018). Air pollution and grassroots echoes of "ecological civilization" in rural China. *China Quarterly*, 320–339.
- Howell, D., Moffatt, S., Prince, H., Bush, J., and Dunn, C.E. (2010). Urban air quality in north-east England: exploring the influences on local views and perceptions. *Risk Anal.* 22, 121–130. <http://data.stats.gov.cn/easyquery.htm?cn=C01>.
- Hunter, L.M., Hatch, A., and Johnson, A. (2004). Cross-national gender variation in environmental behaviors. *Soc. Sci. Q.* 85, 677–694.
- Jasemzadeh, M., Jaafarzadeh, N., Khafae, M.A., Malehi, A.S., and Araban, M. (2016). Predictor of pregnant women's self-care behavior against air pollution: an explanation based on the extended parallel process model (EPPM). *Electron. Physician* 8, 2871–2877. <https://doi.org/10.19082/2871>.
- Jiang, J., Meng, T., and Zhang, Q. (2017). From internet to social safety net: the policy consequences of online participation in China. *SSRN Electron. J.* <https://doi.org/10.1111/gove.12391>.
- Koyama, T., and Smith, P.K. (1991). Showing-off behavior of nursery children. *Aggress. Behav.* 17, 1–10.
- Leeson, S.P.T. (2006). Government's response to Hurricane Katrina: a public choice analysis. *Publ. Choice* 127, 55–73.
- Li, F., Liu, Y., and Meng, T. (2019). Discursive strategy of opinion expression and government response in China: text analysis based on online petitions. *Telematics Inf.* 42, 101238.
- Li, X., Chen, H., and Li, Y. (2020). The effect of air pollution on children's migration with parents: evidence from China. *Environ. Sci. Pollut. Control Ser.* 27, 12499–12513. <https://doi.org/10.1007/s11356-020-07831-3>.
- Liang, Z. (2014). Thinking and research on square dancing and fitness in haze weather. *Business*, 58–59.
- Liu, C.E., and Zhao, B. (2017). Who speaks for climate change in China? Evidence from Weibo. *Climatic Change* 140, 413–422.
- Liu, W., Lai, C.H., and Xu, W. (2018). Tweeting about emergency: a semantic network analysis of government organizations' social media messaging during Hurricane Harvey. *Publ. Relat. Rev.* 44.
- Lu, Y., Wang, Y., Zuo, J., Jiang, H., and Huang, D. (2018). Characteristics of public concern on haze in China and its relationship with air quality in urban areas. *Sci. Total Environ.*
- Marien, S., Hooghe, M., and Quintelier, E. (2010). Inequalities in non-institutionalised forms of political participation: a multi-level analysis of 25 countries. *Polit. Stud.* 58. <https://doi.org/10.1111/j.1467-9248.2009.00801.x>.
- Meng, T., Ping, Y., and Zheng, S. (2015). Public opinion and local fiscal decision making in authoritarian China: based on survey experiment to local government. *J. Public Management*.
- Meng, T., Pan, J., and Yang, P. (2017). Conditional receptivity to citizen participation. *Comp. Polit. Stud.* <https://doi.org/10.1177/0010414014556212>.
- Mo, J., and Hart, S. (2015). Polarized frames on "climate change" and "global warming" across countries and states: evidence from Twitter big data. *Global Environ. Change* 32, 11–17.
- Pantavou, K., Psiloglou, B., Lykoudis, S., Mavrakīs, A., and Nikolopoulos, G.K. (2018). Perceived air quality and particulate matter pollution based on field survey data during a winter period. *Int. J. Biometeorol.* 62, 2139–2150. <https://doi.org/10.1007/s00484-018-1614-3>.
- Reed, A.E., Chan, L., and Mikels, J.A. (2014). Meta-analysis of the age-related positivity effect: age differences in preferences for positive over negative information. *Psychol. Aging* 29, 1–15. <https://doi.org/10.1037/a0035194>.
- Saksena, S. (2012). Public perceptions of urban air pollution risks. *Risk Hazards Crisis Publ. Pol.* 2, 1–19. <https://doi.org/10.1007/s10840-012-9663-8>.
- Sass, V., Kravitz-Wirtz, N., Karceski, S.M., Hajat, A., Crowder, K., and Takeuchi, D. (2017). The effects of air pollution on individual psychological distress. *Health Place*, 72–79.
- Schusky, J. (1966). Public awareness and concern with air pollution in the St. Louis metropolitan area. *J. Air Pollut. Control Assoc.* 16, 72–76. <https://doi.org/10.1080/00022470.1966.10468444>.
- Shi, D., and Chen, S.M. (2019). Web users' attention and government investments in pollution control: an analysis based on big data. *Contemp. Finance Econ.*
- Shin, M., Werner, A.K., Strosnider, H., Hines, L.B., Balluz, L., and Yip, F.Y. (2019). Public perceptions of environmental public health risks in the United States. *Int. J. Environ. Res. Publ. Health* 16. <https://doi.org/10.3390/ijerph16061045>.

- Siamak, Z., Manuel, J., Batista-Foguet, L., and Wunnik, V. (2018). Exploring the public's willingness to reduce air pollution and greenhouse gas emissions from private road transport in Catalonia. *Sci. Total Environ.* 646, 850–861.
- Slovic, P. (1987). Perception of risk. *Science* 236, 280–285. <https://doi.org/10.1126/science.3563507>.
- Tang, X., Zhou, S., Zhao, X., and Wang, Y. (2020). Central-local Relations and Government Responsive Spillover: An Empirical Research of Chinese Citizens' Satisfaction with Environmental Governance. *[J]. Management world* 36, 120–134+249. <https://doi.org/10.19744/j.cnki.11-1235/f.2020.0090.1235/f.2020.0090>.
- Tran, H., Kim, J., Kim, D., Choi, M., and Choi, M. (2018). Impact of air pollution on cause-specific mortality in Korea: results from Bayesian model averaging and principal component regression approaches. *Sci. Total Environ.* 636, 1020–1031. <https://doi.org/10.1016/j.scitotenv.2018.04.273>.
- Tvinnereim, E., Liu, X., and Jamelske, E.M. (2017). Public perceptions of air pollution and climate change: different manifestations, similar causes, and concerns. *Climatic Change* 140, 399–412.
- WHO (2019). Ten Threats to Global Health in 2019 (WHO).
- W J, W. (2017). Does "Hide haze real estate" investment fire? *Money Manag.* 58–59.
- Wang, Z., Ye, X., and Tsou, M.H. (2016). Spatial, temporal, and content analysis of Twitter for wildfire hazards. *Nat. Hazards* 83, 523–540.
- Wang, L., and Watanabe, T. (2019). Effects of environmental policy on public risk perceptions of haze in Tianjin City: a difference-in-differences analysis. *Renew. Sustain. Energy Rev.* 109, 199.
- Wang, J., Obradovich, N., and Zheng, S. (2020). A 43-million-person investigation into weather and expressed sentiment in a changing climate. *One Earth* 2, 568–577.
- Weibo, S. (2019). Weibo Reports Fourth Quarter and Fiscal Year 2018 Unaudited Financial Results.
- Woodside, A.G. (2012). Economic psychology and fashion marketing theory appraising Veblen's theory of conspicuous consumption. *J. Glob. Fash. Mark.* 3, 55–60.
- Wooldridge, J.M. (2009). *Introductory econometrics: a modern approach* (Tsinghua University Press).
- Xu, J., Chi, C.S.F., and Zhu, K. (2018). Concern or apathy: the attitude of the public toward urban air pollution. *J. Risk Res.* 1–17.
- Xue, T., Zhu, T., Zheng, Y., and Zhang, Q. (2019). Declines in mental health associated with air pollution and temperature variability in China. *Nat. Commun.* 10, 2165.
- Yuan, F., and Liu, R. (2018). Crowdsourcing for forensic disaster investigations: Hurricane Harvey case study. *Nat. Hazards* 93, 1529–1546.
- Zhang, and Xiaowen. (2017). The reemerging concern over air pollution in China: the smog of the state's efforts to guide public opinion. *J. Chin. Political Sci.* 23, 519–536.
- Zhang, S., Li, Y., Hao, Y., and Zhang, Y. (2018). Does public opinion affect air quality? Evidence based on the monthly data of 109 prefecture-level cities in China. *Energy Pol.* 116, 299–311.
- Zhang, M., Liu, X., Sun, X., and Wang, W. (2020). The influence of multiple environmental regulations on haze pollution: evidence from China. *Atmos. Pollut. Res.* 11.
- Zheng, S., Wang, J., Sun, C., Zhang, X., and Kahn, M.E. (2019). Air pollution lowers Chinese urbanites' expressed happiness on social media. *Nat. Human Behav.* 3, 237–243. <https://doi.org/10.1038/s41562-018-0521-2>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Public concern	Sina weibo	https://weibo.com/
PM2.5 concentrations	China National Environmental Monitoring Centre (CNEMC)	http://www.cnemc.cn/
Other data	National Bureau of Statistics	http://www.stats.gov.cn
Experimental models:		
LDA topic model	(Tran et al., 2018)	https://doi.org/10.1016/j.scitotenv.2018.04.273
Fixed effect model	(Wooldridge, 2009)	https://doi.org/
Software and algorithms		
Stata	https://www.stata.com/	16MP
Python	https://www.python.org/	3.4.2

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dabo Guan (guandabo@hotmail.com).

Material availability

This study did not generate new unique materials.

Data and code availability

Sina Weibo is the owner and intellectual property rights holder of all information contents that we utilized in this study. We have the right to use the data for the research but are restricted to share the original raw data. All data and codes generated during the study are deposited at Mendeley Data: <https://data.mendeley.com/datasets/bfwnhjkpnm/draft?a=9890309b-b61f-4006-924f-d09253d67cd4>.

METHOD DETAILS

Theoretical framework

In this study, the theory of risk perception, the TPB and government responsiveness are employed to create a conceptual model (Figure 3). The first two theories link people's perceived risk of air pollution with their behaviour of expressing their attitudes and interest demands on social media platforms. In addition, to assess public concerns in driving air quality improvement, the theory of government responsiveness is introduced to analyse the specific mechanism.

Specifically, according to risk perception theory, the public has subjective cognition of various objective risks brought about by air pollution. The experience gained from intuitive judgement and subjective understanding and the public's perception of these risks affect their attitude (Slovic, 1987). Combined with the TPB, the public's attitude and perceived behaviour control (the perception of the difficulty of implementing behaviour) affects people's behavioural intention (Saksena, 2012). Behavioural intention provides motivation for people to express their interest demands through social media. According to the theory of government responsiveness, the government responds to people's demands and pays more attention to air pollution to improve it.

In this paper, we focus on three groups, namely, the young group, the elderly group and the group with women aged between 20 and 35 years. These three groups are often more sensitive to the psychological, physical health and financial risks caused by air pollution (Li et al., 2020; Shin et al., 2019). They have a higher risk perception level of air pollution and stronger motivations to express their own interests

(Sass et al., 2017). While individuals' participation in public debates is resource-demanding, the development of social media provides a convenient, fast and low-cost channel through which vulnerable groups can express their own interests. The perceived behavioural control ability of these groups is stronger, and thus, they are more likely to express themselves on social media platforms. According to the theory of government responsiveness, the government responds to people's environmental demands (Li et al., 2019) and attaches importance to air pollution, which may lead to air quality improvement. The theoretical framework of this paper is shown below.

The daily air quality index and PM_{2.5} concentration released by the Ministry of Ecological Environment of the People's Republic of China are used to measure the severity of air pollution. Perceived risk is the perceived adverse effects of air pollution, including health risk, psychological risk and financial risk. Attitude represents a positive or negative evaluation of behaviour in response to air pollution. In this study, attitude can be denoted as the motivation and initiative of the public to express their own interests. Perceived behavioural control is the perception of the difficulty of implementing behaviour. The development of social media has increased people's perceived behavioural control. In response to air pollution, the public has a variety of behavioural intentions, such as population migration, appeal for governance, and self-protection (for details, please refer to the topic mining analysis in the manuscript). People express their environmental demands on social media platforms to attract government attention. Government concern indicates that the government attaches great importance to air pollution, thus possibly alleviating it altogether.

The air pollution theme identification based on the LDA model

We employed the Latent Dirichlet Allocation (LDA) model to identify air pollution themes. LDA is a probabilistic topic model based on bags of words and generation model, which is a natural language processing algorithm designed to discover hidden topic structures in large-scale documents. This method can simplify the information contained in large-scale unstructured text and transform the text information into digital information that is easy to model. The main premise of the probabilistic topic model is that the document is the probability distribution of the topic, and the topic is the probability distribution of the word. The main feature of LDA is that the document-topic distribution and theme-word distribution are assumed to be random in the process of generating model derivation. Based on the prior knowledge of the Dirichlet distribution as the subject distribution information, the random generation model is solved by a Bayesian method, so that the probabilistic subject model can better describe the document generation process. After the theme number is determined, the text-theme distribution and theme-word distribution are obtained through the LDA model for subsequent modelling of text data.

Specifically, this paper employed Spark Cluster Methods to train the text theme model. First, through the original data pre-processing by data cleaning, word segmentation, and construction of stop word list, all the haze-related posts from January 2017 to September 21, 2018 were taken as the input sample of the model. Then, the perplexity parameter was set to measure the rationality of the theme number. The theme number used in the analysis was selected according to the perplexity and the actual clarity of the theme meaning. Finally, the main outputs of our LDA model were post-theme distribution and theme-word distribution. Due to the large amount of post data, we used Spark Cluster Methods to train the LDA topic model.

Experiments for identifying the emotional level of microblogs in different populations

In order to accurately evaluate the emotional state of different groups in haze weather, we extracted part of microblog data from the obtained data set and analysed their emotional state. The data set was firstly de-duplicated according to the micro-bloggers to prevent repeated selection of the microblogs sent by the same person, and then 100 microblogs were extracted from each group for analysis.

After obtaining microblog samples, the emotional value of each microblog is determined by the evaluation from experts. The emotional evaluation was set with five-score scale from -2 to 2, where evaluation score -2 refers to the microblog in an extremely emotion with negative attitude, and evaluation score 2 refers to the microblog in an extremely emotion with positive attitude. In order to make the scoring results fully credible, we invited 6 volunteers and give them a training for emotional evaluation. Then they were then separately grade the emotional level of each microblog, and the average score were taken as the final emotional score of the microblog. (See [Figure S2](#) for the distribution of emotions)

Definition and measurement of air pollution notability

To explore public concern regarding air pollution in specific groups under different PM_{2.5} concentrations, this paper defines the air pollution notability for women aged between 20 and 35 years, as well as young and elderly individuals. Women aged between 20 and 35 years account for 85.6% of China's pregnant population. The young group is the group aged under 18 years. Elderly individuals are defined as being above 60 years old.

In this paper, air pollution notability is defined as the level of haze-related concern of a specific social group (it can be a social group of a country or region or a social group of a city) that pays attention to air pollution on social networks under a specific concentration of haze within a specific time range. Moreover, air pollution notability is the proportion of the number of haze-related posts on social networks under a specific PM_{2.5} concentration over the total number of posts in that same PM_{2.5} concentration. Air pollution notability is inevitably and logically related to specific social groups and specific time ranges and can be measured as HAI_{gtd} :

$$HAI_{gtd} = \frac{b_{gtd}}{B_{gtd}} \times 10^6 \quad (\text{Equation 1})$$

where b_{gtd} represents the total number of haze-related posts on social networks by a specific social group, g , within a specific time range, t , under a specific PM_{2.5} concentration, d . B_{gtd} represents the total number of posts on social networks by a specific social group, g , within a specific time range, t , under a specific PM_{2.5} concentration, d . We multiply the first part of Equation 1 by 1,000,000 to indicate the number of haze-related posts out of one million common posts and take posts/million as the unit.

The air pollution themes and corresponding theme number involved in Sina Weibo are trained by the LDA model. Public concern in the air pollution theme is measured as the proportion of a certain group's posts on a certain air pollution theme in the collection of haze-related posts under a specific PM_{2.5} concentration within a specific time range; it can be measured as $HTAI_{gtdn}$:

$$HTAI_{gtdn} = \frac{VHT_{gtdn}}{b_{gtd}} \quad (\text{Equation 2})$$

where VHT_{gtdn} represents the total number of haze-related posts of group g on theme n under a specific PM_{2.5} concentration, d , in time period t . VHT_{itdn} can be measured as follows:

$$VHT_{gtdn} = \sum_{j \in bs_{td}^g} \text{sign}(n), \quad (\text{Equation 3})$$

$$\text{where } \text{sign}(n) = \begin{cases} 1, & \text{if } \max(P_j) = p_{jn} \\ 0, & \text{else} \end{cases}$$

where bs_{td}^g represents the set of all haze-related posts for group g at PM_{2.5} concentration d within the time period t and P_j represents the probability distribution of the air pollution theme in the j^{th} post. p_{jn} represents the probability distribution value of topic n in the j^{th} post.

Effect of public concern on air quality improvement

Cities have been grouped according to the number of pollution months they have experienced. Based on the number of months with PM_{2.5} ≥ 75 exposure, cities are divided into 3 pollution levels (pollution month ≤ 3, cities experience a low air pollution level; 3 < pollution month ≤ 6, cities experience a medium air pollution level; pollution month > 6, cities experience a high air pollution level) to explore the disparities in the effects of improving air quality across different cities. The regression model is shown in Equations 4 and 5:

$$\ln(\text{HAZE}_{t+n,d}) = \varphi_1 \ln(\text{HAI}_{t,d}) + \beta_1 \ln(X_{t,d}) + \beta_2 \ln(Z_{t+n,d}) + \gamma_d + \xi_q + \varepsilon_{t,d} \quad (\text{Equation 4})$$

$$\ln(\text{HAZE}_{t+n,d}) = \varphi_1 \ln(\text{HAI}_{t,d}) + \varphi_2 \ln(\text{HAI}_{t,d}) * G_d + \beta_1 \ln(X_{t,d}) + \beta_2 \ln(Z_{t+n,d}) + \gamma_d + \xi_q + \varepsilon_{t,d} \quad (\text{Equation 5})$$

where $HAZE_{t+n,d}$ represents the air pollution concentration of city d in month $t + n$. n represents the lag time. $HAI_{t,d}$ represents the public notability level. G_d represents cities with different pollution levels. $X_{t,d}$ represents a group of social and economic variables for month t in city d . The socioeconomic variables include the income level ($INCOME_{t,d}$) and industry structure ($Industry_{t,d}$). $Z_{t+n,d}$ represents a group of natural environment variables in month $t+n$ in city d . The natural environment variables include average wind speed ($Wind_{t+n,d}$) and rainfall situation ($Rain_{t+n,d}$). These data are matched with the local weather statistics. The time trend is controlled through the seasonal dummy variable ζ_q . γ_d is the individual effect of city d , which controls for all time-invariant individual characteristics, like national haze control policies, which have not changed much since the implementation of the Action Plan of Air Pollution Prevention and Control in 2013. $\varepsilon_{t,d}$ represents random error. Standard errors are clustered at the city quarterly level (Wooldridge, 2009).

Robustness check

$$\ln(HAZE_{t+n,d}) = \varphi_1 \ln(HAI_{t,d}) + \varphi_2 \ln(HAI_{t,d}) * N_d + \beta_1 \ln(X_{t,d}) + \beta_2 \ln(Z_{t+n,d}) + \gamma_d + \xi_q + \varepsilon_{t,d} \quad (\text{Equation 6})$$

where $HAZE_{t+n,d}$ represents the air pollution concentration of city d in month $t+n$. n represents the lag time. N_d represents the regions of northern and southern China. $\varepsilon_{t,d}$ represents random error. Standard errors are clustered at the city quarterly level.

$$\ln(HAZE_{t+n,d}) = \varphi_1 \ln(HAI_{t,d_young}) + \varphi_2 \ln(HAI_{t,d_aged}) + \varphi_3 \ln(HAI_{t,d_exmother}) + \beta_1 \ln(X_{t,d}) + \beta_2 \ln(Z_{t+n,d}) + \gamma_d + \xi_q + \varepsilon_{t,d} \quad (\text{Equation 7})$$

where $HAZE_{t+n,d}$ represents the air pollution concentration of city d in month $t+n$. n represents the lag time. HAI_{t,d_young} , $HAI_{t,d_exmother}$ and HAI_{t,d_aged} represent the air pollution notability level of young people, women aged between 20 and 35 years and elderly people in month t in city d , respectively. $\varepsilon_{t,d}$ represents random error. Standard errors are clustered at the city quarterly level.

Mechanism analysis

How can the public notability level improve air quality? Based on a causal-step approach, we analyse the mechanism by which the air pollution notability level affects air quality from the perspective of government response. The regression model is shown in Equations 8, 9 and 10:

$$\ln(HAZE_{t+n,d}) = \varphi_1 \ln(HAI_{t,d}) + \beta_1 \ln(X_{t,d}) + \beta_2 \ln(Z_{t+n,d}) + \gamma_d + \xi_q + \varepsilon_{t,d} \quad (\text{Equation 8})$$

$$\ln(RES_{t,d_government}) = \varphi_1 \ln(HAI_{t,d}) + \beta_1 \ln(X_{t,d}) + \beta_2 \ln(Z_{t,d}) + \gamma_d + \xi_q + \varepsilon_{t,d} \quad (\text{Equation 9})$$

$$\ln(HAZE_{t+n,d}) = \delta \ln(RES_{t,d_government}) + \varphi_1 \ln(HAI_{t,d}) + \beta_1 \ln(X_{t,d}) + \beta_2 \ln(Z_{t+n,d}) + \gamma_d + \xi_q + \varepsilon_{t,d} \quad (\text{Equation 10})$$

where $RES_{t,d_government}$ represents the response of the local government, which comes from the government's responsive blog posts on Weibo. γ_d represents the individual effect of the city. The time trend is controlled through the seasonal dummy variable ζ_q . $\varepsilon_{t,d}$ represents random error. Standard errors clustered at the city quarterly level.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data source of air pollution public concern and air pollution

The data on public concern about air pollution are from Sina Weibo. We use "haze" as a search keyword to extract the air pollution posts from the full sample of 3.5 billion posts from January 2017 to September 2018. The dataset of this study contains demographic information and air-pollution-related posts by the public on social media platforms from 329 cities in China (see Table S3 and Figure S1 for details). Nearly 3.8 million haze-related posts are extracted as a data pool, which includes the time, location and content of each post, as well as the age, gender and phone model of each blogger.

The data of PM_{2.5} concentrations are from the *China National Environmental Monitoring Centre* (CNEMC). There are 1,436 sites for air quality monitoring, covering 366 cities across China. We collect daily air quality data across 322 cities from the CNEMC and match them with the haze-related post data according to time and location. We then build a mapping table regarding PM_{2.5} concentrations and post attributes of Sina Weibo (see [Tables S1](#) and [S2](#) for details).