AEA Papers and Proceedings 2022, 112: 151–155
https://doi.org/10.1257/pandp.20221069

PERCEPTION OF POLICIES AND POLARIZATION

Scapegoating during Crises†

By Leonardo Bursztyn, Georgy Egorov, Ingar Haaland, Aakaash Rao, and Christopher Roth*

Economic crises are often accompanied by waves of antiminority behavior (Allport 1954; Staub 1989). The Weimar Republic, followed by the German Reich, passed increasingly repressive laws targeting Jews as the country grappled with the Great Depression (Doerr et al. 2021). In Rwanda, a collapse in the price of coffee was a major cause of the 1994 genocide (Newbury 1995). In the United States, Donald Trump won the presidency on a platform blaming immigrants for middle-class stagnation and, during the early stages of the COVID-19 pandemic, blaming immigrant groups for spreading the virus (Bartoš et al. 2021).

Why does economic hardship so often precipitate antiminority behavior? One explanation is that the frustration and sense of injustice ignited by economic crises lead people to seek out "someone to blame" (Bauer et al. 2021). Moreover, opportunistic politicians can often exploit crises by supplying persuasive antiminority narratives (Voigtländer and Voth 2015).

In this article, we build upon the framework developed in Bursztyn et al. (2022) to propose an additional mechanism: crises can provide intolerant people with a plausible *rationale* for their views, increasing their willingness to engage in antiminority behavior by lowering the expected social sanctions from doing so. By this logic, crises increase antiminority behavior not only by changing people's attitudes toward minorities, but also by making them more willing to express preexisting prejudice. This is consistent with a recent body of evidence (Cantoni, Hagemeister, and Westcott 2019; Fisman, Hamao, and Wang 2014; Fouka and Voth 2021) that finds that latent historical antipathy toward a group can be "activated" by crises or political opportunists.

A simple example captures the intuition. Consider a xenophobe who dislikes immigrants due to a distaste for foreign cultures but cannot express this motive without incurring social sanctions. Despite widespread stereotypes about immigrants "stealing jobs" and depressing wages for low-skilled native workers (Haaland and Roth 2020), it is hard to claim genuine concern when unemployment is low and wages are increasing. Yet during an economic crisis, concerns about immigrants' effects on the labor market are far more credible—particularly when these concerns are stoked by charismatic political entrepreneurs. Observers judging the motives underlying antiminority behavior now face a signal extraction problem: the behavior may be driven by innate xenophobia, but it also may be driven by genuine concerns about losing one's job, being unable to provide for one's family, etc. Xenophobes can thus *pool* with people with such genuine concerns, enabling them to engage in public antiminority behavior at a lower social cost. An important implication is the existence of a "social amplifier," as described in Bursztyn et al. (2022): if the crisis leads some people to adopt more antiminority positions due to genuine concerns about, for instance, losing their job, then xenophobes face lower social costs for

* Bursztyn: University of Chicago and NBER (email: bursztyn@uchicago.edu); Egorov: Kellogg School of Management and NBER (email: g-egorov@kellogg.northwestern.edu); Haaland: University of Bergen and CESifo (email: Ingar.Haaland@uib.no); Rao: Harvard University (email: arao@g.harvard.edu); Roth: University of Cologne and CEPR (email: roth@wiso.uni-koeln.de). Roth acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC 2126/1-390838866.

† Go to https://doi.org/10.1257/pandp.20221069 to visit the article page for additional materials and author disclosure statement(s).

expressing antiminority positions and are thus more willing to do so.

This paper presents an experiment examining how economic crises affect social inference about the motives underlying xenophobic behavior.

I. Experimental Design and Sample

We collected data from a large and heterogeneous sample of 1,952 American respondents in December 2021 and January 2022 in collaboration with Prolific, a survey provider commonly used in economic experiments (Haaland, Roth, and Wohlfart, forthcoming). We aimed to recruit approximately equal numbers of Joe Biden and Trump voters; 51.1 percent of our respondents voted for Biden, with the remaining 48.9 percent voting for Trump. The mean age in our sample is 37.9 years, 49.6 percent of our respondents are male, 82.3 percent of our respondents are White, and 66.1 percent of our respondents have a college degree. Following a set of background questions, we provide respondents with a vignette about "Mike," a blue-collar worker who enjoyed a stable and well-paying manufacturing job prior to the 2008 financial crisis. During the financial crisis, Mike's factory suddenly went bankrupt and Mike lost his job.

Respondents are randomized into two treatments: *Before Crisis* (971 respondents) and *After Crisis* (981 respondents). In both treatments, Mike joins an anti-immigration organization. In the *Before Crisis* treatment, Mike joined the anti-immigration organization "shortly before the financial crisis," whereas in the *After Crisis* treatment, Mike joined the organization "shortly after the financial crisis." This treatment variation thus cleanly manipulates whether Mike had a plausible *rationale* for joining the anti-immigrant organization, holding other potential confounds fixed. In particular, if Mike joined before the financial crisis, his decision clearly must have been motivated by factors other than the crisis and his resulting unemployment.

Following the approach developed in Bursztyn et al. (2022), we measure beliefs about Mike's motivations for joining the anti-immigrant organization by asking respondents the following open-ended question: "Why do you think Mike joined this organization? 2–3 sentences should be enough." As a more natural elicitation than a structured belief measure, this approach avoids priming respondents about any particular dimensions and allows us to observe what comes to people's minds when they learn about Mike's decision to join the anti-immigrant organization.

II. Results

Our analysis begins with a simple word-counting procedure. After preprocessing the text data, we create two indicator variables to capture respondents' inferences about Mike's motives. The first indicator takes the value one if a response contains any of the following xenophobia-related stems: *xenophob*, *racis*, *intoler*, *bias*, and *bigot*. The second indicator takes value one if the respondent uses any of the following labor-related stems: *labor*, *job*, *unemploy*, and *work*. Across conditions, 77 percent of respondents mention labor-related terms, while 7.4 percent of respondents mention xenophobia-related terms. While Trump and Biden voters are about equally likely to mention labor market concerns (mentioned by 77.3 percent of Trump voters and 76.9 percent of Biden voters), Biden voters are much more likely to use xenophobia-related terms to describe why Mike joined the organization (mentioned by 12.2 percent of Biden voters compared to only 2.3 percent of Trump voters).

Panel A of Figure 1 displays treatment effects on xenophobia-related terms. In line with the intuition discussed above, respondents are much more likely to characterize Mike using xenophobia-related terms when he joined the anti-immigration organization before, rather than after, the crisis: 9.9 percent of respondents in the *Before Crisis* treatment mention racism-related terms in the open-ended responses, compared to only 4.9 percent of respondents in the *After Crisis* treatment ($p < 0.001$). Turning to the second indicator on labor-related terms, panel B of Figure 1 shows that respondents are also more likely to ascribe Mike's anti-immigrant behavior to concerns about the labor market in the *After Crisis* treatment: 73.8 percent of respondents in the *Before Crisis* treatment mention labor-related terms in the open-ended responses, compared to 80.2 percent of respondents in the *After Crisis* treatment ($p < 0.001$). Thus, the treatment appears to induce near one-to-one substitution between xenophobia-related terms and labor-related terms.

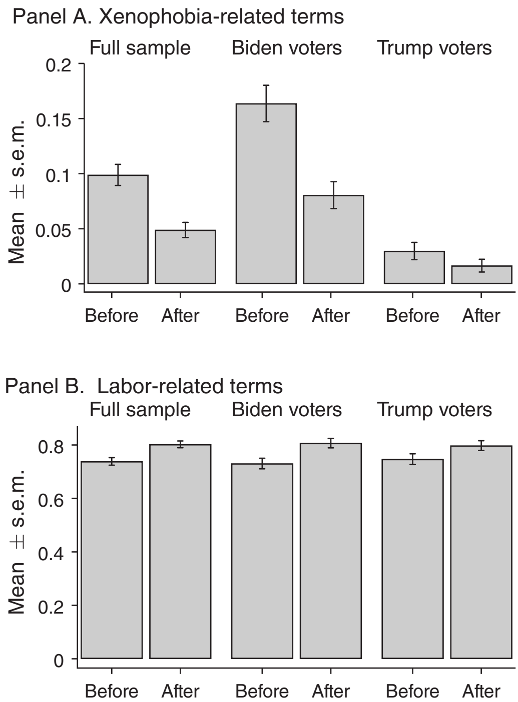

Figure 1. Treatment Effects on Inference about Motives

Notes: The figure shows the fraction of respondents using any xenophobia-related terms (panel A) and any labor-related terms (panel B) separately by treatment condition for the full sample ($n = 1{,}952$) as well as separately for Biden ($n = 998$) and Trump voters ($n = 954$). The error bars indicate the standard error of the mean.

We also find substantial heterogeneity between Trump and Biden voters in treatment effects on the use of xenophobia-related terms. While Biden voters are 8.3 percentage points less likely to mention xenophobia-related terms in the *After Crisis* treatment (a 50.7 percent decrease relative to the 16.4 percent of Biden voters who mention these terms in the *Before Crisis* treatment, $p < 0.001$), Trump voters are only 1.3 percentage points less likely to mention xenophobia-related terms in the *After Crisis* treatment (a 44.2 percent decrease relative to the 3 percent of Trump voters who mention these terms in the *Before Crisis* treatment, $p = 0.175$). This heterogeneity is statistically significant ($p = 0.003$). In contrast, there is no significant treatment effect heterogeneity between Trump and Biden voters when examining the use of labor-related terms ($p = 0.501$).

We now turn to a less structured approach to measuring how our treatment shifted respondents' perceptions of the motives underlying Mike's decision. A common approach to measuring differences in open-ended text across groups is to examine how predictive text is of treatment group status: the more predictive is the text, the larger the between-group differences. We implement this approach using a two-stage classifier. In the first stage, we use Bidirectional Encoder Representations from Transformers (Devlin et al. 2018), a state-of-the-art natural language processing technique that constructs high-dimensional vector representations of text responses capturing semantic meaning. In the second stage, we use a neural network to predict treatment status based on these high-dimensional vectors. We train our classifier on 80 percent of the data and calculate model accuracy using the remaining 20 percent.

As shown in row 1 of Table 1, we find that our model is 69 percent accurate in predicting respondents' treatment status based on their open-ended response. This is substantially better than chance ($p < 0.001$), confirming that our treatment variation indeed induces significant differences in respondents' perceptions of Mike's motives. To benchmark the extent to which the treatment induces differential perceptions against the effect of different demographic characteristics, we repeat this exercise, predicting various binary demographic variables (whether the respondent self-identifies as a liberal, has a four-year college degree, is above median age, etc.) using the open-ended responses and reporting accuracies and associated p-values in Table 1. Strikingly, the model is almost equally accurate in predicting the treatment condition as it is in predicting whether the respondent was a Biden (versus Trump) voter in 2020, and it exhibits greater accuracy (relative to the base rate in the population) in predicting treatment status than any other characteristic. The fact that the effect of the crisis on responses is quantitatively large relative to that of other demographic characteristics suggests that crises may have strong effects on the interpretation, and by extension the incidence, of antiminority behavior across heterogeneous contexts.

A drawback of this procedure is that the classifier is a black box: it is challenging to understand precisely what factors have predictive power and thus what dimensions the treatment

Table 1—Classifier Accuracy

Dimension	Accuracy	Rate	p-value
After crisis	0.69	0.51	$<$ 0.001
Biden voter	0.71	0.55	$<$ 0.001
College	0.62	0.62	$>$ 0.99
High income	0.64	0.63	0.915
Male	0.52	0.51	0.681
White	0.70	0.79	$<$ 0.001
Old	0.59	0.51	0.001

Notes: Table displays the accuracy of the classifier in predicting each characteristic based on respondents' text responses. "Rate" is the base rate of the characteristic in the population. "High income" and "old" are indicators for whether the respondent's characteristic is greater than or equal to the median in the sample. p-value is calculated from a t-test of whether the accuracy is equal to the base rate.

is shifting. One approach is to manually inspect (out-of-sample) responses with the highest predicted probabilities of belonging to either condition. The results are largely consistent with our analyses above: The response with the highest predicted probability of belonging to the *After Crisis* treatment is

> I'm sure Mike saw or read unsavory coverage that said there was an influx of immigrants coming to America and we didn't have the infrastructure to support it. He needed a person to blame for his job loss.

The response with the highest predicted probability of belonging to the *Before Crisis* treatment is

> He probably felt that immigrants would take his job. Either that or he's most likely racist or prejudiced.

To more systematically probe differences in responses across conditions, we examine which words, or phrases of up to three words, are most characteristic of either condition. We follow the approach of Gentzkow and Shapiro (2010) to calculate the χ^2 statistic for each phrase, where a higher statistic indicates that a phrase is more characteristic of a given condition. Figure 2 plots the top 100 phrases by their χ^2 statistic, with positive values corresponding to phrases more characteristic of the *After Crisis* condition and negative values to phrases more characteristic of the *Before Crisis* crisis condition. Consistent with our results above, we find that respondents in the *After Crisis* condition are more likely to use phrases relating to Mike's job loss ("losing

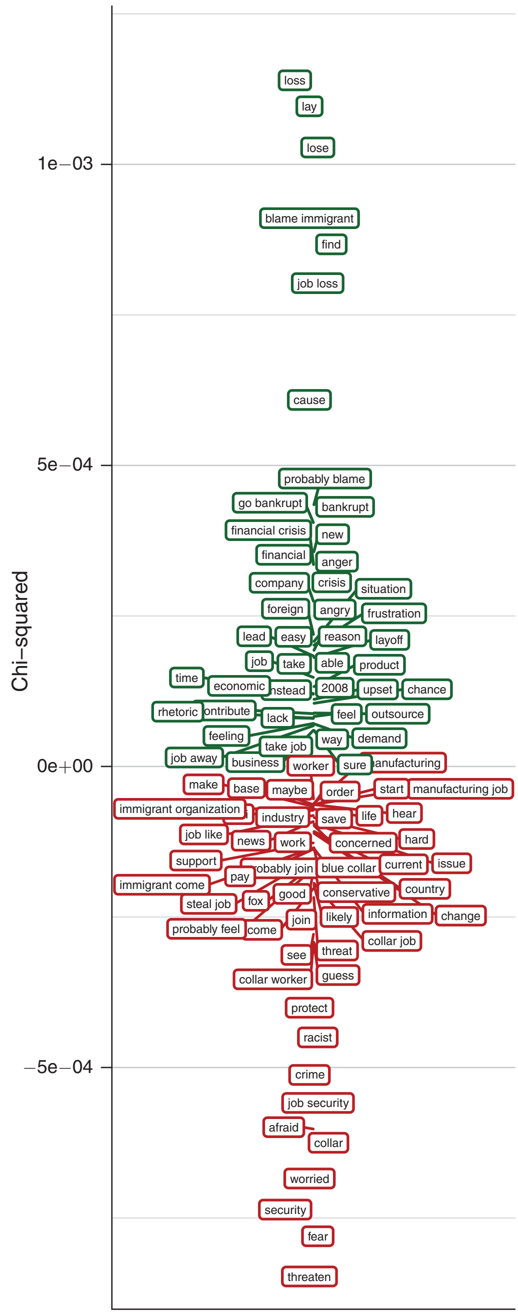

Figure 2. Most Characteristic Phrases of Each Condition

Notes: Figure displays the phrases with the 100 largest χ^2 statistics. Figure omits the word "blame," which has a χ^2 statistic of 0.00207, in order to better scale the other phrases. Phrases with a positive χ^2 statistic are more characteristic of the *After Crisis* condition; phrases with a negative χ^2 statistic are more characteristic of the *Before Crisis* condition.

his job," "went bankrupt") and the resulting emotions ("upset"), whereas respondents in the *Before Crisis* condition use terms either referencing Mike's *fear* of losing his job in the future or relating to his underlying type ("racist," "conservative," "ignorant," "Fox").

III. Conclusion

Antiminority behavior is often stigmatized, but economic crises can facilitate scapegoating: downturns shift social inference about the motives underlying antiminority behavior, reducing the associated social costs. Our results suggest several promising avenues for future research. First, to what extent can crises serve as coordination devices, facilitating mass expressions of hostility toward minorities in areas where such prejudice was previously latent? This may be particularly relevant in settings in which xenophobes underestimate the share of other people who share their views (e.g., Bursztyn, Egorov, and Fiorin 2022), where crises and resulting expression may help correct these misperceptions and further lower the social cost of antiminority behavior. Second, what characteristics of a given area or minority group affect the extent to which crises can unleash antiminority behavior? For example, how does the visibility of minority groups shape their exposure to scapegoating? Third, can disseminating positive rationales about immigrants—for example, providing research evidence about their overall positive impact on the economy—make scapegoating during a crisis less socially acceptable?

REFERENCES

Allport, Gordon W. 1954. *The Nature of Prejudice*. Cambridge, MA: Addison-Wesley.

Bartoš, Vojtěch, Michal Bauer, Jana Cahlíková, and Julie Chytilová. 2021. "COVID-19 Crisis and Hostility against Foreigners." *European Economic Review* 137: Article 103818.

Bauer, Michal, Jana Cahlíková, Julie Chytilová, Gérard Roland, and Tomas Zelinsky. 2021. "Shifting Punishment on Minorities: Experimental Evidence of Scapegoating." NBER Working Paper 29157.

Bursztyn, Leonardo, Georgy Egorov, and Stefano Fiorin. 2020. "From Extreme to Mainstream: The Erosion of Social Norms." *American Economic Review* 110 (11): 3522–48.

Bursztyn, Leonardo, Georgy Egorov, Ingar K. Haaland, Aakaash Rao, and Christopher P. Roth. 2022. "Justifying Dissent." NBER Working Paper 29730.

Cantoni, Davide, Felix Hagemeister, and Mark Westcott. 2019. "Persistence and Activation of Right-Wing Political Ideology." Collaborative Research Center TRR 190 Discussion Paper 143.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." https://arxiv.org/pdf/1810.04805.pdf.

Doerr, Sebastian, Stefan Gissler, Jose-Luis Peydro, and Hans-Joachim Voth. 2021. "Financial Crises and Political Radicalization: How Failing Banks Paved Hitler's Path to Power." Bank for International Settlements Working Paper 978.

Fisman, Raymond, Yasushi Hamao, and Yongxiang Wang. 2014. "Nationalism and Economic Exchange: Evidence from Shocks to Sino-Japanese Relations." *Review of Financial Studies* 27 (9): 2626–60.

Fouka, Vasiliki, and Hans-Joachim Voth. 2021. "Collective Remembrance and Private Choice: German-Greek Conflict and Consumer Behavior in Times of Crisis." https://vfouka.people.stanford.edu/sites/g/files/sbiybj4871/f/collective_memory_0.pdf.

Gentzkow, Matthew, and Jesse M. Shapiro. 2010. "What Drives Media Slant? Evidence From U.S. Daily Newspapers." *Econometrica* 78 (1): 35–71.

Haaland, Ingar, and Christopher Roth. 2020. "Labor Market Concerns and Support for Immigration." *Journal of Public Economics* 191: Article 104256.

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart. Forthcoming. "Designing Information Provision Experiments." *Journal of Economic Literature*.

Newbury, Catharine. 1995. "Background to Genocide: Rwanda." *Issue: A Journal of Opinion* 23 (2): 12–17.

Staub, Ervin. 1989. *The Roots of Evil: The Origins of Genocide and Other Group Violence.* New York: Cambridge University Press.

Voigtländer, Nico, and Hans-Joachim Voth. 2015. "Nazi Indoctrination and Anti-Semitic Beliefs in Germany." *Proceedings of the National Academy of Sciences* 112 (26): 7931–36.