The Dynamics of Discrimination and Assimilation: Theory and Evidence

Gianandrea Lanzara
Bocconi University

Sara Lazzaroni
University of Bologna

Paolo Masella
University of Bologna

Mara P. Squicciarini
Bocconi University *

July 20, 2023

Abstract

This paper provides a unified framework to study the interactions between a minority group, a majority group, and political leaders. Theoretically, we set up a dynamic model of discrimination and assimilation choices and characterize the effect of a shock to the return to discriminatory actions. Empirically, we use text-analysis techniques on a novel dataset of tweets of "White American" and Chinese users from January to August 2020. We show that anti-Chinese discrimination increased following the COVID-19 outbreak and Trump's tweet referring to COVID-19 as "Chinese virus." In response, Chinese users tweeted assimilation-related content and criticism against the Chinese Communist Party.

JEL-Classification: J7, Z1, H8
Keywords: discrimination of minorities, culture, social media

*We thank Enrico Cantoni, Simon Görlach, Uta Schönberg, Carlo Schwarz, Vincenzo Scrutinio, and seminar participants at Bocconi University, at the Second Text as Data Workshop, and at the 2022 Winter Meeting of the Econometric Society for useful comments and suggestions. We thank Caterina Alfonzo, Ludovica Massacesi, and Federico Scabbia for excellent assistance throughout the construction of the dataset. Financial support from the PRIN, Grant # PRIN 2017ATLJHB, is gratefully acknowledged.

1 Introduction

Discrimination against minorities is a long-standing phenomenon, and the spreading of antiminority sentiments is still very common today. Given the consequences of discrimination on the social, economic, and political spheres, the literature has largely investigated the origins of antiminority sentiments. Economic and cultural forces, social media, and the behavior of political leaders have been considered among the possible drivers (see Dustmann and Preston (2001); Mayda (2006); Becker and Pascali (2019); Müller and Schwarz (2020, 2021), among others). However, little is known about how minorities react to discrimination, in particular when it comes from prominent political leaders; we do not know whether they respond by isolating themselves from the majority or emphasizing their identification with the country of residence. Our goal in this paper is to provide a unified setting to study, from both a theoretical and an empirical perspective, the interactions between members of a discriminated minority, members of the majority group, and political leaders.

We view discrimination and assimilation decisions as part of a dynamic process in which individuals build their own identities (Bénabou and Tirole, 2011). Such decisions are not made in isolation; rather, they take place within a social environment that is shaped by the identities of the other members of society (Bisin et al., 2011), including prominent individuals, such as leaders (Prummer and Siedlarek, 2017; Verdier and Zenou, 2018).

To represent these features theoretically, we set up a dynamic discrete-choice model of discrimination and assimilation choices. Individuals are forward-looking, and their actions influence the evolution of their identities, where identity is modeled as a form of capital - discrimination capital, for the majority group, and assimilation capital, for the minority group. The incentives to discriminate and assimilate depend on the distribution of discrimination capital in the majority group (the "discriminatory environment") and on the actions of a leader. Two features make the characterization of the equilibrium challenging in this setup. First, because the leader's behavior cannot be perfectly anticipated, the leader's actions are a source of uncertainty. Second, because the leader is aware that her actions influence the evolution of the discriminatory environment, we have to account for her strategic interaction with the majority group. To prove the existence of an equilibrium, we combine tools from the literature on large, dynamic economies (Jovanovic and Rosenthal, 1988; Miao, 2006) with the game-theoretical literature on supermodular games (Van Zandt and

Vives, 2007).

Within this framework, we study the social consequences of (1) an exogenous change on the return to discrimination, and (2) a discriminatory action on the part of the leader. Employing the supermodularity properties of the problem, together with the tools of robust comparative statics analysis (Milgrom and Shannon, 1994; Acemoglu and Jensen, 2015), we analytically characterize the impact of both types of shocks on the share of agents in the majority group who decide to discriminate and on the share of agents in the minority group who decide to assimilate at each time period. Regarding discrimination,our framework predicts that the share of discriminating users increases after the exogenous circumstances become more favorable to discrimination, and after the leader discriminates. Regarding assimilation, our framework clarifies that the reaction of the minority group depends on whether the incentive to assimilate is stronger or weaker in a more discriminatory environment and with a discriminatory leader. This is ultimately an empirical question. These comparative statics incorporate the inter-temporal considerations of the individuals, the general equilibrium effect arising from the collective behavior of the group, and the strategic effects from the leader's future probability to discriminate. .

We then bring the model to the data, focusing on the diffusion of anti-Chinese sentiments in the United States following the COVID-19 crisis. Given the association of the virus with the location where it was initially discovered (China's Wuhan Province), we interpret the novel coronavirus outbreak as an exogenous shock to the return to discrimination against the Chinese minority. This allows us to identify the causal effect of the health shock on the discrimination decisions of the majority against Chinese people (and China, more broadly), and on the assimilation behavior of such minority.[1] In this context, the use of pandemic-related anti-Chinese and anti-Asian rhetoric by prominent political leaders (e.g., former U.S. President Trump) provides us with a unique setting in which to investigate the role of political leaders in further exacerbating health-related discrimination shocks.

[1]The medical sociology literature has emphasized the ability of epidemic outbreaks to trigger antiminority sentiments and xenophobia as "[o]utbreaks create fear, and fear is a key ingredient for racism and xenophobia to thrive" (Devakumar et al., 2020). Antiminority sentiments have been historically linked to epidemics due to ideological scapegoating—for example, during the Black Death, when Jews were accused of well-poisoning in the context of antichrist conspiracy theories (Voigtländer and Voth, 2012; Jedwab et al., 2019)—or due to social stigmatizing of people and places associated with the birth and spread of a virus, as in the case of sites affected by plague, cholera, and yellow fever during colonial times, and, more recently, as in the case of Wuhan and China with COVID-19 (White, 2020).

To investigate health-related and leader-induced discriminatory and assimilation behaviors, we analyze the social media activity of a large sample of Twitter users in the United States. We focus on two key dates: (1) March 9, 2020, when several restrictive measures were adopted in the United States, such as isolation and quarantine for suspected and confirmed cases, cancelling of public events and suspension of in-person classes at universities and schools (some district- and state-level closures soon followed); (2) March 17, 2020, when then–U.S. President Donald Trump tweeted the phrase "Chinese Virus." In line with the building blocks of the model, the majority group is represented by the sample of "White (non-Hispanic) American" users ("White" or "White American" users henceforth), and the minority group is represented by the sample of Chinese users residing in the United States (the "Chinese minority" or "members of the Chinese community" henceforth). As described in detail in Section 3 and Appendix D, we use users' self-descriptions contained in their Twitter profiles to obtain information on their group identification.

The advantage of using social media data is that it allows us to build several high-frequency measures of both discriminatory and assimilation attitudes. Past empirical studies on discrimination and assimilation typically relied on either survey responses (see, among others, Aspachs-Bracons et al., 2008; Manning and Roy, 2010), census records (Fouka, 2019; Abramitzky et al., 2020; Fouka, 2020a), or on evidence from experimental manipulation of perceptions of the discriminatory context in the lab (see Bertrand and Duflo, 2017, for a review of experiments on discrimination). Using supervised and unsupervised text-analysis techniques, we are able to study *actual* individual-level discrimination activities of White Americans and assimilation attitudes of the members of the Chinese community that clearly reflect responses to the shocks, rather than being equilibrium outcomes determined by changes in behavior by natives and by minorities (e.g., marriage decisions or decisions to participate to the labor market). To measure discrimination, we track the use of racial slurs and hateful discourse toward the targeted group (as in, for example, Lu and Sheng, 2020; Tahmasbi et al., 2021). To measure assimilation behavior, we move one step forward with respect to the economics literature by providing three *original measures* of assimilation that build on the work of sociologists studying the assimilation experience of second-generation Chinese/Asian individuals. In particular, we rely on two indicators based on the stated national (beyond ethnic) identity and on one indicator of disidentification from the original ethnic group.

Exploiting the timing of the health and leader shocks, we show, by means of a trend-break model and a regression-discontinuity design, that both shocks generated a sharp increase in the share of posted discriminatory text. Using the lenses of our model, both March 9 and March 17, 2020, represent positive shocks to the returns to discrimination. Because China is the place of origin of the virus and of the Chinese community, the cost to discriminate against an individual of Chinese origin decreases; similarly, discriminating against the Chinese community for the virus is a way to hold the minority group responsible for the shock. These effects are ultimately amplified by the subsequent adoption of discriminatory rhetoric by the leading political figure. In terms of magnitudes, we find a 2.25-percentage-point increase in the daily probability of posting a tweet containing abusive language against the Chinese minority after March 9, followed by an 11.69-percentage-point increase after March 17.

Finally, we turn to the question of whether increased discrimination raises or reduces the return to assimilation. In line with the smaller increase in discrimination on March 9, the regression-discontinuity results point to a nonsignificant reaction by Chinese Twitter users on March 9. However, after the larger increase in discrimination on March 17, we find a significant increase in tweets with assimilation (and disidentification) content: the daily probability of posting a tweet with assimilation content increases by 2.13 percentage points. The richness of the data allows us to show that the most integrated members of the Chinese community (that is, Twitter users with a higher share of "friends" located in the United States) were more strongly asserting their Chinese American identity. At the same time, both the most and the least integrated users (definitions based on share of "friends" in the United States) are equally likely to distance themselves from the Chinese Communist Party (CCP).

Our results are robust to considering specific subsets of users for both White and Chinese users, thus alleviating the concern of selection issues (see the robustnesses of Sections 5.1 and 5.2). Moreover, our results remain robust when alternative measures of assimilation are considered.

Altogether, these findings suggest that, following the political leader's discriminatory behavior, both the majority and minority group discriminated against react, respectively, by increasing discrimination and by attempting to counter the new discrimination wave by asserting one's belonging to the prevalent group or distancing oneself from their original ethnic group.

Literature. While most studies analyze the issues of discrimination against and integration of minorities separately (see below for specific references), this paper is the first, to the best of our knowledge, to offer a unified framework to study, from both a theoretical and an empirical perspective, the interactions between a majority group, a discriminated minority, and political leaders.

On the theory side, our model builds on the economic literature that incorporates identity into models of economic decisions. Following Akerlof and Kranton (2000)'s seminal paper, we view discrimination (for the majority group) and assimilation (for the minority group) as rational choices whose payoffs depend on individual identity, as well as on the identities of other members of society. At the same time, like in Bénabou and Tirole (2011) and Bisin et al. (2011), identity is an endogenous outcome of the agents' decisions; in particular, we follow Bénabou and Tirole (2011) in modeling identity as a form of capital that individuals can build up through their actions. An important difference between our approach and these studies is that we set up an infinite-horizon model with a continuum of forward-looking individuals, rather than assuming myopic behavior or stylized time structures.

Thus, from a methodological point of view, our model is most closely related to the literature on anonymous sequential games (Jovanovic and Rosenthal, 1988), and to the macroeconomic literature on dynamic economies with a large number of agents (Acemoglu and Jensen, 2015), in particular to the subbranch that considers the role of aggregate shocks (Bergin and Bernhardt, 1992, 1995; Miao, 2006). To the best of our knowledge, we are the first to apply this class of models to address questions of discrimination against and assimilation by minorities. A novel element of our framework is that aggregate shocks will emerge as a natural consequence of the optimal behavior of a rational agent, i.e., the leader, who exerts direct influence on individual payoffs, rather than the outcome of an exogenous stochastic process.

In this regard, our framework is also related to the recent literature that introduces rational leaders into models of cultural dynamics (Prummer and Siedlarek, 2017; Verdier and Zenou, 2018). In these studies, the leader is an agent with a specific objective function and choice set, different from the "ordinary" members of his group. By contrast, in our model, the leader is a member of her group in all respects, except that her identity directly matters for payoffs, whereas other members' identities only enter payoffs via their aggregate distribution, not individually.

Finally, our theoretical analysis is related to the literature on robust comparative statics (Milgrom and Roberts, 1994; Milgrom and Shannon, 1994; Acemoglu and Jensen, 2015). In particular, Acemoglu and Jensen (2015) present comparative results for the stationary equilibria of large, dynamic economies in response to a permanent change in the value of an exogeneous parameter. Motivated by our empirical application, we build on their approach to obtain results for the effect of a *temporary* exogenous shock on the set of *sequential* equilibria in a nonstationary environment.

Meanwhile, from an empirical point of view, by analyzing the assimilation attitudes of a minority group facing discrimination, this paper contributes to a growing literature studying the cultural and social integration of immigrants. Building on seminal works showing that the cultural identity of immigrants significantly explains variation in several socioeconomic outcomes (Fernández and Fogli, 2009; Giuliano, 2007; Bisin et al., 2008), researchers have begun investigating the assimilation decisions of immigrant minorities. Recent studies focus on the choices of first names for children (Fouka, 2019; Abramitzky et al., 2020; Fouka, 2020a), hosting region/countries' language adoption and proficiency (Bleakley and Chin, 2010; Avitabile et al., 2013), intermarriage patterns (Gould and Klor, 2016a; Bisin and Tura, 2019; Fouka, 2020a; Guirkinger et al., 2021; Fouka et al., 2022), and self-reported national identity (Aspachs-Bracons et al., 2008; Manning and Roy, 2010; Clots-Figueras and Masella, 2013; Abdelgadir and Fouka, 2020). These studies largely exploit the introduction of specific immigration policies and reforms (such as compulsory language laws, citizenship laws, no-fault divorce laws, and veil bans) to achieve causal identification.

Our contribution to this literature is twofold. First, in this paper, we specifically consider how minorities react to changes in the discriminatory environment driven by the behavior of a prominent political leader. Second, we are able to show causal evidence on *both* the discrimination attitudes of the majority *and* the assimilation reactions of the minority discriminated against.

Showing evidence on both discrimination and assimilation of the minority discriminated against in the context of an epidemic allows us to make a step forward with respect to studies showing increasing antiminority sentiments after economic shocks (Anderson et al., 2017; Becker and Pascali, 2019; Anderson et al., 2020; Grosfeld et al., 2020) and epidemics (Jedwab et al., 2019), including COVID-19 (Lu and Sheng, 2020; Ziems et al., 2020; Dipoppa et al., 2021; Tahmasbi et al., 2021).

A recent growing literature links traditional and new media (including social media) to the spread of violence against minorities[2]. Regarding traditional media, DellaVigna et al. (2014),

[2]Other studies have looked at the relationship between social media and information diffusion (Halberstam and

Yanagizawa-Drott (2014), and Adena et al. (2015) show that exposure to radio propaganda can contribute to ethnic hatred and violence. Similarly, in the context of new media, Bursztyn et al. (2019) and Müller and Schwarz (2021) study how social media such as Facebook and VK can foster hatred of minorities. By contrast, Bailey et al. (2022) exploits Facebook data to build assimilation measures for Syrian refugees in Germany. Differently from these studies, in this paper we exploit an exogenous trigger of possible discriminatory behavior to show how different social groups respond in terms of discrimination and assimilation behavior.

Finally, an emerging literature studies how leaders are able to influence the behavior of the population at large by affecting political preferences and mobilizing people in social movements (Cagé et al., 2020; Dippel and Heblich, 2021), and by influencing the societal perception of social norms regarding discrimination (Bursztyn et al., 2020; Grosjean et al., 2020; Müller and Schwarz, 2020). In line with the findings of Barberá et al. (2019), which suggest that leaders generally follow the issue priorities set by the public, we model the leader as a more important player within the majority, with a behavior that amplifies the effects of a given discriminatory shock. However, in contrast with other studies, we show that different groups react differently to the leader's actions.

The rest of this paper proceeds as follows. Section 2 presents the theoretical model. Section 3 discusses the background and data for our empirical analysis. Section 4 illustrates the empirical strategy. Sections 5 and 6 present our main findings. Section 7 concludes.

2 Theory

We set up a dynamic discrete-choice model with forward-looking individuals who belong to one of two social groups: the majority or the minority. The dynamic incentives in the model stem from a distinction between actions and identity: the agent's choices contribute to shaping her future identity, while her identity influences the payoff for each action. In this view, identity is a form of capital that agents build through their choices. For example, capital may represent the stock or intensity of a majority (e.g., White American) identity for both social groups. While the majority group accretes this identity through actions that mark its difference from the minority group (e.g., discriminatory actions), the minority group, conversely, acts in ways that emphasize its adherence

Knight, 2016), corruption (Enikolopov et al., 2018), and protests Acemoglu et al. (2017)

to the majoritarian identity (e.g., assimilation actions).[3] A second key feature of our framework is that the payoff for an action depends on the identities of all other members of a group. For instance, from the majority group's perspective, the return to a discriminatory action depends on the discriminatory environment, defined as the distribution of discrimination capital across all agents in the majority group. Finally, we allow for the presence of a leader, i.e., a member of the majority group who has a large influence on society. The leader is forward-looking and understands that her actions will affect the evolution of the discriminatory environment. We use this framework to study how an exogenous shock to the return to the discrimination, such as the COVID-19 outbreak in the United States, affects the discrimination and assimilation choices of individuals in each group, including the leader, as well as their identities.[4]

2.1 Setup

The majority. At each time t, an agent from the majority group selects an action d from the set $\{0, 1\}$, where $d = 1$ stands for discriminating and $d = 0$ stands for not discriminating. Her choice gives her the following flow payoff:

$$\tilde{u} : \{0, 1\} \times \mathcal{K} \times \mathcal{E} \times \mathcal{P} \times \Theta \times \mathcal{S} \to \mathbb{R}$$

This specification for the flow-payoff function incorporates five determinants of discriminatory behavior. First, $K \in \mathcal{K}$ denotes the discrimination capital of an agent, which we interpret as a notion of individual identity. Second, $\varepsilon \in \mathcal{E}$ represent a vector of choice-specific idiosyncratic shocks to the discrimination payoff. Together, $\mathcal{K}$ and $\mathcal{E}$ represent the space of individual characteristics. Third, $p \in \mathcal{P}$ denotes a probability measure over $\mathcal{K}$. Thus, p summarizes the distribution of individual identities across agents in the majority group. By including it in the payoff function, we allow the behavior of individual agents to depend on the discriminatory environment prevailing in the society. Fourth, $\theta \in \Theta$ denotes the identity of a leader. As will be clear later, agents are not able

[3] An alternative interpretation is that capital represents the intensity of the own-group identity for both groups. For instance, if one defined the minority group along ethnic lines, then its capital would reflect the identity of its origin country or culture. The act of assimilating should, then, be viewed as a disinvestment, i.e., a costly distancing from the origin-country identity. In this case, the assumptions we will make later regarding the complementarity between the action and other variables must be modified accordingly.

[4] In what follows, all partial orders are denoted with $\geq$. Probability measures are ordered according to first-order stochastic dominance. The integral operator denotes the Lebesgue integral. Discrete sets are endowed with the discrete topology. Sets of probability measures are endowed with the topology of weak convergence.

to perfectly forecast the behavior of the leader, therefore θ is a source of aggregate uncertainty in the model. Fifth, $s \in \mathcal{S}$ is a parameter that captures the exogenous factors related to the return of discrimination. Agents are assumed to have perfect foresight of the future path $(s_1, s_2, \dots)$, and to discount the future with discount rate β.

Assumption 1 (Domain sets)**.** *The set $\mathcal{K}$ is a compact subset of $\mathbb{R}$.*

Assumption 2 (Discount factor)**.** $0 < \beta < 1$.

Assumption 3 (Payoff function)**.** *The payoff function $\tilde{u}$ is*

A: bounded, jointly continuous, and measurable.

B: supermodular.

C: weakly increasing in $K \in \mathcal{K}$.

D: additive in the idiosyncratic shocks, i.e., $\tilde{u}(d, \cdot, \varepsilon) = u(d, \cdot) + \varepsilon(d)$.

Assumptions 1–3.A are standard assumptions that ensure that the discrimination problem has a well-behaved recursive formulation (we list further technical assumptions in Appendix A). Assumptions 3.B and 3.C are the key assumptions of the model. Assumption 3.B implies that the discrimination decisions are positively related to the aggregate discrimination behavior of the majority group and the exogenous circumstances. If Assumption 3.C also holds, then discrimination decisions also increase with the level of individual discrimination capital. This assumption captures, in reduced form, all factors and motivations that sustain discrimination actions as rational behavior.[5] Assumption 3.D is standard in the literature on dynamic discrete-choice modeling (Rust, 1988; Aguirregabiria and Mira, 2007), as it provides a simpler way to solve the Bellman equation in the recursive problem.

Example 1. Suppose that discrimination capital can take two values, high (K^h) or low (K^l). Also, suppose that $\Theta = \mathcal{K}$. Clearly $\mathcal{K} = \{K^l, K^h\}$ is compact. Fix a partial history θ^t for the leader, and let $p \in \mathcal{P}$ be the capital distribution at time t after this history. To illustrate, suppose that the

[5] According to social identity theory, for instance, outgroup discrimination is a means to reinforce ingroup identity; this in turn may have psychological or economic benefits (Shayo, 2020)

capital distribution of the majority group and the leader's identity enter the payoff function via a single aggregate, D, defined from

$$D = (1-\mu)\int_{\mathcal{K}} xp(\mathrm{d}x) + \mu\theta$$

where μ and $1-\mu$ denote, respectively, the weights of the leader and the majority group. Thus, suppose that

$$u(d, K, \theta, p, s) = F(K) - C(d, D, s)$$

where F is increasing in K and C is increasing in d. In this formulation, the payoff u can be decomposed in a return component, which depends on discrimination capital, and in a cost component, which depends on the action, on the discrimination index D, and on the exogenous circumstances. Assumption 3.A is satisfied if F and C are bounded and C is continuous in D and s. Assumption 3.B is satisfied if it is relatively less costly to discriminate when D is higher, either because θ is higher, or because the distribution of discrimination capital in the majority group is higher (in the stochastic dominance sense).

If an individual with discrimination capital K chooses an action d, her discrimination capital for the next period is drawn from a probability distribution $Q(K, d; \cdot)$.[6] Three aspects of the transition rule Q are worth mentioning. First, the value of discrimination capital in the next period depends on the current value of capital. This captures the fact that individual identities exhibit some persistence. Second, Q also depends on the agent's choice. Therefore, agents can actively shape their future identities through their choices. Third, the identity formation process may include a stochastic component.[7]

Assumption 4 (Stochastic kernel)**.** *Q is*

A: increasing in d, that is: $Q(K, 1; \cdot) \succsim Q(K, 0; \cdot)$ for any $K \in \mathcal{K}$.

B: increasing in K, that is: $Q(K_2, d; \cdot) \succsim Q(K_1, d; \cdot)$ for any $K_2, K_1 \in \mathcal{K}$ such that $K_2 \geq K_1$, and for any $d \in \{0, 1\}$.

[6]Formally, $Q : \mathcal{K} \times \{0, 1\} \times \mathcal{B}(\mathcal{K}) \to [0, 1]$ is a transition kernel, where $\mathcal{B}(\mathcal{K})$ denotes the Borel field of the set $\mathcal{K}$. As an example, $Q(K, d; A)$ is interpreted as the probability that discrimination capital in the next period lies in the set $A \subset \mathcal{B}(\mathcal{K})$, given the current value of capital K and the discrimination choice d.

[7]This formulation encompasses the deterministic case as the special case, wherein the transition kernel is degenerate.

C: stochastically supermodular in (K, d)*, i.e.,*

$$\int_{\mathcal{K}} f(x)Q(K,1;\mathrm{d}x) - \int_{\mathcal{K}} f(x)Q(K,0;\mathrm{d}x)$$

is weakly increasing in K *on* $\mathcal{K}$*, for any increasing function* $f : \mathcal{K} \to \mathbb{R}$*.*

Assumption 4.A is saying that discrimination leads to the accumulation of discrimination capital in the next period. Assumption 4.B also naturally interprets that discrimination capital in the next period tends to be higher when the current value of capital is high, and that this is true given any discrimination choice. Assumption 4.C is saying, loosely speaking, that discrimination is more effective at increasing the likelihood of high values of capital in the next period, when current capital is already high.

Example 2. When $\mathcal{K}$ is finite, the transition kernel Q takes the form of two transition matrices, one for $d = 1$ and one for $d = 0$. Suppose the transition matrices are given by

d=1

	K^h	K^l
K^h	π^1	$1 - \pi^1$
K^l	π^2	$1 - \pi^2$

d=0

	K^h	K^l
K^h	π^3	$1 - \pi^3$
K^l	π^4	$1 - \pi^4$

where $0 < \pi^i < 1$, $i = 1 \ldots 4$. Assumption 4.A requires $\pi^1 \geq \pi^3$ and $\pi^2 \geq \pi^4$. Assumption 4.B requires: $\pi^1 \geq \pi^2$ and $\pi^3 \geq \pi^4$. Assumption 4.C requires: $\pi^1 - \pi^3 \geq \pi^2 - \pi^4$. These conditions are satisfied, for instance, for $\pi^1 = 1$, $\pi^2 = \pi^3 = 1/2$, and $\pi^4 = 0$.

Finally, the idiosyncratic shocks are drawn from a time-invariant density λ.

The minority. At each time t, an agent from the minority group selects an action a_t from the set $\{0, 1\}$, where $a = 1$ stands for assimilating and $a = 0$ stands for not assimilating. Her choice gives her the following flow payoff:

$$\tilde{u}^a : \{0,1\} \times \mathcal{K}^a \times \mathcal{E}^a \times \mathcal{P}^a \times \Theta \times \mathcal{P} \to \mathbb{R}.$$

The flow-payoff function for the minority group is analogous to the one assumed for the majority group. The individual return to assimilation may depend on the individual assimilation capital

$K^a \in \mathcal{K}^a$, on an idiosyncratic component $\varepsilon^a \in \mathcal{E}^a$, on the distribution of assimilation capital across other individuals in the minority group $p^a \in \mathcal{P}^a$, on the discriminatory environment $p \in \mathcal{P}$, and, finally, on the leader's identity $\theta \in \Theta$. If an individual in the minority group with assimilation capital K^a chooses an action a, she draws her next-period value of assimilation capital from a probability distribution $Q^a(K^a, a; \cdot)$ (see footnote 6), whereas the idiosyncratic shock is drawn from the invariant distribution λ^a.

To prove the existence of an equilibrium for the minority group, we will need the following assumption on the payoff function.

Assumption 5 (Payoff function)**.** *The payoff function* $\tilde{u}^a$

A: is bounded, jointly continuous, and measurable.

B: exhibits increasing differences pairwise in a, K^a*, and* p^a*.*

C: weakly increasing in $K^a \in \mathcal{K}^a$*.*

D: additive in the idisyncratic shocks, i.e., $\tilde{u}^a(a, \cdot, \varepsilon^a) = u(a, \cdot) + \varepsilon^a(a)$*.*

The assumption that the assimilation variables (the action, the individual identity, and the identity of other members of the minority group) reinforce each other is symmetric to Assumption 3.B on the corresponding variables for the majority group. The difference here is that we are not taking a stand on the relationship between the assimilation variables and the other aspects of the minority group's problem, that is, the discriminatory environment and the leader's (discriminatory) identity. We also impose assumptions on Q^a that parallel Assumption 4 for the majority group, with a similar interpretation. Also, we assume that the two groups have the same discount factor.

The leader. A leader is an agent whose actions have a large influence on society. Ordinary individuals are too "small" to affect aggregate outcomes, and their identities enter the payoff function only via their aggregate distribution (the discrimination environment). Instead, the leader's identity exerts a distinct influence on the majority group's incentives to discriminate. Furthermore, since the leader is rational, she is aware of it. Apart from this key difference, the leader has the same objectives and the same choice set as ordinary individuals. However, because she considers the impact of her actions on the social environment, we will need the following assumption on the leader's payoff:

Assumption 6. *The leader's payoff function u^ℓ is weakly increasing in p.*

For the purpose of simplifying the exposition, we will also introduce an additional restriction for the leader, namely, that her identity can take values in the finite set.

Assumption 7. *The set Θ is finite, and the leader's transition function Q^ℓ is defined on $\Theta \times \{0,1\} \times \mathcal{B}(\Theta)$.*

This assumption implies that the set of all possible histories of the leader's identities is countable and that our theoretical arguments can be developed pointwise (i.e., history by history). [8]

2.2 Timing

Time starts at $t = 0$ with given capital distributions p_0 and p_0^a and leader's identity $\theta_0 \in \Theta$. At the beginning of period t, all members of the society observe their own identity and the value of their private idiosyncratic shock, the distribution of identities across the majority and minority groups, the identity of the leader, and the current value of the exogenous circumstances $s_t \in \mathcal{S}$. Based on this information, the members of the majority group and the leader decide whether to discriminate, and the members of the minority group decide whether to assimilate. Then, flow payoffs are received. At the end of the period, all identities update: individuals in the majority group draw a new value of discrimination capital from Q, individuals in the minority group draw a new value of assimilation capital from Q^a, and the leader's identity evolves according to Q^ℓ.

2.3 Uncertainty

Ordinary individuals face two sources of uncertainty in the model. The first is related to their individual draws from Q and λ (for the majority group) and from Q^a and λ^a (for the minority group). In what follows, we assume that all these distributions satisfy a no-aggregate-uncertainty condition (Jovanovic and Rosenthal, 1988). Loosely speaking, this means that when a continuum of individuals draws from Q, the resulting distribution of discrimination capital in the next period is deterministic.

[8]See Bergin and Bernhardt (1995) for a treatment of aggregate uncertainty with Θ uncountable.

The second source of uncertainty is related to the leader's behavior. Ordinary individuals are unable to perfectly forecast the leader's actions (and therefore her identity in the future), because they do not observe the realization of her idiosyncratic shock. Since the leader's identity influences the decisions of all members of society, the future paths of the capital distributions also become stochastic. In other words, the leader's behavior introduces aggregate uncertainty in the model.

Let $\bar{p}$ and $\bar{p}^a$ denote infinite *sequences* of capital distributions, that is $\bar{p} = \{p_t\}_{t=0}^{\infty}$ and $\bar{p}^a = \{p_t^a\}_{t=0}^{\infty}$, respectively, with $p_t \in \mathcal{P}$ and $p_t^a \in \mathcal{P}^a$ for all $t = 0, 1 \ldots$. Technically, aggregate uncertainty is difficult to deal with because $\bar{p}$ and $\bar{p}^a$ become random variables and cannot be taken as given in the individual optimization problem. As the previous discussion suggests, however, the sequences $\bar{p}$ and $\bar{p}^a$ become deterministic once the sequence of the leader's identities is conditioned on. To put it differently, if the future path of the leader's identities was known up to some period t, then individuals would be able to correctly anticipate the capital distributions at that date. This *conditional* no-aggregate-uncertainty condition (Bergin and Bernhardt, 1992) can be used to replace the sequences $\bar{p}$ and $\bar{p}^a$ with sequences of mappings from partial histories to capital distributions in the individual problem.

We now introduce some notation to formalize this notion. Let $\theta^t = (\theta_0, \theta_1, \theta_2, \ldots, \theta_t)$ denote a partial history of the leader's identities to time t, and let $\Theta^t = \times_{\tau=0}^{t}\Theta$ for $t = 1, 2, \ldots$ be the set of all possible partial histories to time t. Conditional no-aggregate uncertainty implies that it is possible to define a *function* $\pi_t : \Theta^{t-1} \to \mathcal{P}$ that maps partial histories into probability measures over $\mathcal{K}$. Let $\mathcal{F}^t$ denote the space of all such functions for each t. Then $\bar{\pi} = \{\pi_t\}_{t=1}^{\infty}$, such that each π_t is viewed as an element of $\mathcal{F}^t$, is a deterministic object from the point of view of ordinary individuals. We write $\mathcal{F}^{\infty} = \times_{t=1}^{\infty} F^t$ to denote the set where sequences $\bar{\pi}$ lie. To clarify, consider the following example.

Example 3. Suppose the leader's identity takes two values, K^l and K^h, so that $\Theta = \{K^l, K^h\}$, and let $K_0 \in \Theta$ denote her initial identity. At time 0, the capital distribution is observed. At time 1, it depends on the leader's identity at time 0; since this is also observed, π_1 is trivial. At the time of making decisions in period 2, previous events may have in general given place to two partial histories: $\theta_1^1 = \{K_0, K^l\}$ and $\theta_2^1 = \{K_0, K^h\}$. The set of possible partial histories observed at time 2 is therefore $\Theta^1 = \{\theta_1^1, \theta_2^1\}$. Looking into the future from time 1, the capital distribution should be expected to differ at the beginning of period 2, because everyone conditions their choices on θ^1 at time 1. Therefore the function $\pi_2 : \Theta^1 \to \mathcal{P}$ will say, for instance, $\pi_2(\theta_1^1) = p_1$ and $\pi_2(\theta_2^1) = p_2$, where p_1, p_2 are two generic capital distributions in $\mathcal{P}$. Similarly, at time 3 there will be four possible partial histories in Θ^2, and the function π_3 will associate an element of $\mathcal{P}$ to each of them. The sequence of π_t maps for all future time periods is $\bar{\pi} = \{\pi_1, \pi_2, \pi_3, \dots\}$.

In a similar way, we define $\bar{\pi}^a$ as the sequence of maps $\pi_t^a \in \mathcal{F}^{a,t}$ such that $\pi_t^a : \Theta^{t-1} \to \mathcal{P}^a$ for all t, $\bar{\pi}^a$ being an element of $\mathcal{F}^{a,\infty} = \times_{t=1}^{\infty} \mathcal{F}^{a,t}$.

A key difference with respect to other studies where aggregate uncertainty is present is that the "aggregate shocks" are not the realizations of an exogenous stochastic process but rather the outcome of the optimizing behavior of a specific agent, i.e., the leader. Let a *strategy* for the leader be a map $w_t^\ell : \Theta^t \times \mathcal{E} \to \{0, 1\}$ from her state variables (θ^t, ε) into her choice set. Let W_t denote the space of such strategies and, as usual, let $W^\infty = \times_{t=1}^{\infty} W^t$ denote the space of the corresponding sequences $\bar{w}^\ell = \{w_1^\ell, w_2^\ell \dots\}$. The majority and minority groups take $\bar{w}^\ell$ as given when making their choices and rationally form their beliefs on the stochastic evolution of the leader's identity. Because the leader's private shock is not observed, at each time $t = 1, 2, \dots,$ these beliefs can be described by a map q_t from the set of partial histories to probability measures over Θ, with $\bar{q} = \{q_t\}_{t=1}^{\infty}$. In Section 2.5 below, we show how to construct $\bar{q}$.

2.4 The One-Person Decision Problem

The majority group. A discrimination plan prescribes a discrimination action at each future time period and for each possible future contingency. At time 0, an individual selects a discrimination plan to maximize the expected discounted sum of her future payoffs. In so doing, she takes the sequences $\bar{\pi}$ and $\bar{w}^\ell$ as given. To characterize the optimal plan, we formulate the discrimination problem as a nonstationary dynamic programming problem. Specifically, define the policy correspondence from

$$G_t(K, \varepsilon, \theta^t, \bar{\pi}, \bar{w}^\ell, \bar{s}) = \underset{d \in \{0,1\}}{\arg\max}\, u(d, K, \theta_t, \pi_t(\theta^{t-1}), s_t) + \varepsilon(d) + \beta \mathbb{E}_{Q,q,\lambda}[V_{t+1}(K', \varepsilon', \theta^{t+1}, \bar{\pi}, \bar{s})] \quad (1)$$

Here, $\mathbb{E}_{Q,q,\lambda}[V_{t+1}(\cdot)]$ represents the continuation value of the problem, as seen from the perspective of time t. The expectation is taken with respect to the distributions λ, Q, and q_t, reflecting the sources of uncertainty faced by the individual. Equation (1) then describes the optimal discrimination choices for a member of the majority group with discrimination capital K and private value ε, after having observed a history of the leader's identities θ^t. The discrimination choice d affects the instantaneous payoff u as well as the continuation value V_{t+1}, because it contributes to determining, via Q, the identity of the individual in the next period. While her identity in the next period K' does matter for her individual payoff, the ordinary individual understands that it will be irrelevant for the evolution of the capital distribution. Therefore, the aggregate sequence $\bar{\pi}$ is taken as given in (1), and so are the leader's behavior and the sequence of exogenous circumstances $\bar{s}$.

The leader. Let $u^\ell(d, \theta, p, s) = u(d, \theta, \theta, p, s)$. Then the discrimination correspondence for the leader can be written as

$$G_t^\ell(\theta^t, \varepsilon, \bar{\pi}, \bar{s}) = \underset{d \in \{0,1\}}{\arg\max}\, u^\ell(d, \theta_t, \pi_t(\theta^{t-1}), s_t) + \varepsilon(d) + \beta \mathbb{E}_{Q^\ell, \lambda}\left[V_{t+1}^\ell(\varepsilon', \theta^{t+1}, \bar{\pi}, \bar{s})\right] \tag{2}$$

Equation (2) describes the optimal discrimination choices of the leader; it is similar to equation (1) except for one important difference. As for the other members of the majority group, the leader's discrimination choice affects her identity in the next period via Q^ℓ; however, the leader also recognizes that her future identity will affect the evolution of the capital distribution, because it enters as an argument in π_{t+1} (as well as in π_{t+2}, π_{t+3}, etc.), the reason being, again, that the leader's identity affects the discrimination incentives of all members of the majority. As a result, the leader takes into account both the direct effect of her action on her individual payoff (as a member of the majority) and the effect arising from changes in the discriminatory environment. Finally, the leader takes the expectation only with respect to the distributions λ and Q^ℓ.

Recall that a strategy w_t^ℓ for the leader at time t gives the leader's action for given θ^t and ε, while keeping the aggregate behavior of the majority group, $\bar{\pi}$, fixed. The set of strategies that are consistent with the leader's optimal behavior for each $\bar{\pi}$ is the leader's *best response*. Formally, the best-response correspondence α^ℓ is defined from

$$\alpha^\ell(\bar{\pi}, \bar{s}) = \{\bar{w}^\ell : w_t^\ell(\theta^t, \varepsilon) \in G_t^\ell(\theta^t, \varepsilon, \bar{\pi}, \bar{s}) \text{ all } \theta^t \in \Theta^t, \varepsilon \in \mathcal{E}, t = 1, 2, \dots\} \tag{3}$$

The assimilation problem. The assimilation problem for the minority group parallels the discrimination problem for the majority group discussed in Section 2.4 above. The policy correspondence describing the optimal assimilation choices for an individual in the minority group is

$$G_t^a(K^a, \varepsilon^a, \theta^t, \bar{\pi}^a, \bar{w}^\ell, \bar{\pi}) = \arg\max_{a \in \{0,1\}} u(a, K^a, \pi_t^a(\theta^{t-1}), \pi_t(\theta^{t-1})) + \varepsilon^a(a) + \beta \mathbb{E}_{Q^a, q, \lambda^a}[V_{t+1}^a(K^{a'}, \varepsilon^{a'}, \theta^{t+1}, \bar{\pi}^a, \bar{\pi})] \quad (4)$$

2.5 The Law of Motion for Capital

After making their discrimination or assimilation choices, ordinary individuals draw an updated value of capital for the following period. We now discuss how the policy correspondences G_t, G_t^a and the transition kernels Q, Q^a can be used to describe the evolution of the capital distribution.

First, we define the fraction of individuals in the majority group with capital K who choose to discriminate after history θ^t from

$$\hat{g}_t(K, \theta^t, \bar{\pi}, \bar{w}^\ell, \bar{s}) = \lambda(\{\varepsilon \in \mathcal{E} : g_t(K, \varepsilon, \theta^t, \bar{\pi}, \bar{s}) = 1\}), \quad t = 1, 2 \ldots \quad (5)$$

for some sequence of policy functions $\{g_t\}_{t=1}^{\infty}$, such that each g_t is a selection from G_t.[9]

Second, we can compute the probability that a member of the majority group with capital K at time t ends up with capital $K' \in A$ at time $t+1$ as

$$\hat{Q}_t(K; A) = \hat{g}_t(K, \theta^t, \bar{\pi}, \bar{w}^\ell, \bar{s}) Q(K, 1; A) + (1 - \hat{g}_t(K, \theta^t, \bar{\pi}, \bar{w}^\ell, \bar{s})) Q(K, 0; A) \quad (6)$$

for all $A \in \mathcal{B}(\mathcal{K})$. Intuitively, a fraction $\hat{g}_t(K, \cdot)$ will sample an updated value from $Q(K, 1; \cdot)$, while the remaining $1 - g_t(K, \cdot)$ fraction will sample from $Q(K, 0; \cdot)$.

Third, we can integrate (6) over the capital distribution at time t, p_t, to obtain an updated capital distribution for the next period. In equilibrium, this distribution has to be equal to $p_{t+1} = \pi_{t+1}(\theta^t)$: the anticipated sequence $\bar{\pi}$ turns out to be consistent ex post (and this must hold along

[9]Note that $\hat{g}_t$ is uniquely defined, although the policy correspondence G_t is in general multivalued, because individuals will be indifferent between discriminating and not discriminating only on a zero measure set.

all sequences $\bar{\theta} \in \Theta^\infty$). Formally, in equilibrium

$$[\pi_{t+1}(\theta^t)](A) = \int_{\mathcal{K}} \hat{Q}_t(K, ; A)[\pi_t(\theta^{t-1})](\mathrm{d}K) \qquad (7)$$

$$\text{for all } A \in \mathcal{B}(\mathcal{K}), t = 1, 2, \ldots, \text{ all } \bar{\theta} \in \Theta^\infty$$

A similar construction applies to the minority group. In equilibrium,

$$[\pi^a_{t+1}(\theta^t)](A) = \int_{\mathcal{K}} \hat{Q}^a_t(K^a, ; A)[\pi^a_t(\theta^{t-1})](\mathrm{d}K^a), \qquad (8)$$

$$\text{for all } A \in \mathcal{B}(\mathcal{K}^a), t = 1, 2, \ldots, \text{ all } \bar{\theta} \in \Theta^\infty$$

where $\hat{g}^a$ and $\hat{Q}^a$ are defined analogously to (5) and (6), respectively.

The arguments leading to (6) and (7) can also be applied to the leader. Let $\hat{w}^\ell_t$ define the probability that the leader discriminates and $\hat{Q}^\ell_t$ denote the transition kernel for the leader's identity for $t = 1 \ldots$. Meanwhile, $\hat{Q}^\ell_t$ is an objective probability at time t. In an equilibrium with rational agents, it should coincide with the beliefs q_t, that is

$$[q_t(\theta^t)](A) = \hat{w}^\ell_t(\theta^t) Q^\ell(\theta_t, 1; A) + (1 - \hat{w}^\ell_t(\theta^t)) Q^\ell(\theta_t, 0; A) \qquad (9)$$

for all $A \in \mathcal{B}(\mathcal{K})$ and all $t = 1, 2 \ldots$.

2.6 Discrimination Equilibrium

We can now turn to the definition of an equilibrium for the majority group and the leader. Broadly speaking, in a discrimination equilibrium, we require that (i) all individuals behave optimally, given their beliefs on the behavior of other individuals; and (ii) their beliefs are indeed consistent with the optimal behavior of other individuals. Formally, we have the following:

Definition 1. A discrimination equilibrium with a leader is a sequence of policy correspondences $\bar{G}$ for the individuals of the majority group, a sequence of strategies $\bar{w}^\ell$ for the leader, an aggregate sequence $\bar{\pi} \in \mathcal{F}^\infty$, and a sequence of beliefs $\bar{q} \in \mathcal{F}^\infty$ on the leader's discrimination capital, such that

i. $\bar{G}$ is defined from equation (1).

ii. the sequence $\bar{\pi}$ satisfies (7).

iii. the sequence of beliefs $\bar{q}$ is defined from (9).

iv. $\bar{w}^{\ell} \in \alpha^{\ell}(\bar{\pi}, \bar{s})$, where α^{ℓ} is defined in (3).

Parts *(i)–(iii)* of this definition describe an equilibrium "internal " to the majority group, keeping the leader's behavior fixed at a generic sequence of strategies $\bar{w}^{\ell}$. Part *(i)* requires that ordinary individuals behave optimally at each "node" in history, given the sequence $\bar{\pi} \in \mathcal{F}^{\infty}$ and their beliefs about the leader's behavior. Part *(ii)* is a general equilibrium condition: while ordinary individuals take the aggregate behavior of the majority group as given, its aggregate behavior must turn out to be consistent with their individual choices; because the leader's choices are uncertain, the condition must hold pointwise along all histories $\bar{\theta}$. Part *(iii)* requires that the majority group's beliefs are consistent with the leader's given strategy. Then, part *(iv)* also requires that the leader's behavior is optimal.

To prove the existence of an equilibrium, Proposition A.3 in the Appendix first provides an aggregation result for the majority group. By this approach, the majority group can be viewed as a single agent, whose behavior is determined from the consistent aggregation of a continuum of uncoordinated individual choices. In particular, the majority group "plays" a sequence of capital distributions $\bar{\pi} \in \mathcal{F}^{\infty}$ that satisfies conditions (1) and (7). Furthermore, Proposition A.3 also shows that this so-defined "aggregate best response" possesses some desirable properties of best responses in supermodular games. Thanks to these properties, in a second step we study the equilibrium between the leader and the majority group as a strategic game between two players, and we adapt the existence results in Van Zandt and Vives (2007). We then have

Theorem 1. *There exists a discrimination equilibrium with a leader.*

Having proved the existence of an equilibrium, we now characterize some of the empirical implications of the model. The next proposition focuses on how the discrimination behavior of the majority group changes in response to a discriminatory action on part of the leader. In our empirical setting, we interpret Trump's Chinese-virus tweets as a leader's discriminatory action.

Let $E^M(\bar{s})$ in $\mathcal{F}^{\infty} \times \mathcal{H}^{\infty}$ denote the set of equilibrium pairs $(\bar{\pi}, \bar{w}^{\ell})$ that are consistent with Definition 1 for a given $\bar{s} \in \mathcal{S}$. If the equilibrium is unique, i.e., E^M contains only one element, the proposition says that the share of individuals in the majority group who decide to discriminate is expected to increase in the aftermath of the leader's discriminatory action. If there are multiple equilibria, the result holds for the greatest and the least equilibria.

The share of individuals in the majority group who decide to discriminate is defined from

$$\phi_t^d(\theta^t, \bar{\pi}, \bar{w}^\ell, \bar{s}) = \int_{\mathcal{K}} \hat{g}_t(K, \theta^t, \bar{\pi}, \bar{w}^\ell, \bar{s})[\pi_t(\theta^{t-1})](\mathrm{d}K), \text{ for } \theta^t \in \Theta^t, \quad t = 1, 2, \dots \tag{10}$$

where $\hat{g}_t$ has been defined earlier in (5). We then have

Proposition 1. *Let $(\bar{\pi}_\vee, \bar{w}^\ell_\vee)$ denote the greatest equilibrium in E^M, and let $\phi^d_{\vee,t}$ be defined according to (10) for $\bar{\pi}$ equal to $\bar{\pi}_\vee$. Let $d^\ell_t \in \{0,1\}$ denote the observed leader's action at time $t > 0$. Then,*

$$\mathbb{E}[\phi^d_{\vee,t+\tau}|\theta^t, d^\ell_t = 1] \geq \mathbb{E}[\phi^d_{\vee,t+\tau}|\theta^t, d^\ell_t = 0]$$

for all $t > 0$, $\tau > 0$, and all $\theta^t \in \Theta^t$, where the expectation is taken with respect to density of $\theta^{t+\tau}$ conditional on θ^t and d^ℓ_t. The same inequality holds at the least equilibrium $(\bar{\pi}_\wedge, \bar{w}^\ell_\wedge)$ in E^M.

This result is an application of Proposition A.16 in Appendix A.[10] To gain intuition, suppose that the equilibrium is unique. From the perspective of an ordinary individual in the majority group, the incentive to discriminate goes up for three main reasons. First, when the leader discriminates, the likelihood of the leader's identity becoming more discriminatory in the next period increases. Second, in general equilibrium the aggregate behavior of the majority group amplifies this effect, because the distribution of discrimination capital in the majority group shifts upwards. Third, the strategic interaction between the leader and the majority group ensures that these two effects reinforce each other, so that it becomes more likely to observe histories of highly-discriminatory identities in all subsequent periods. However, because the leader also receives idiosyncratic shocks, the result holds in expectation.

2.7 The Assimilation Equilibrium

Because the behavior of the minority group does not affect the leader's payoff, there is no strategic interaction between the minority group and the leader. The minority group takes the pair $(\bar{\pi}, \bar{w}^\ell)$

[10] In the Appendix we also discuss the relevant order on E^M

as the given equilibrium outcome of the interaction between the leader and the majority group.

An assimilation equilibrium is a sequence of policy correspondences $\bar{G}^a$ for the individuals of the minority group and a sequence of mappings $\bar{\pi}^a$ such that: (1) individual behavior is optimal given the behavior of the minority group; (2) the aggregate behavior of the minority group results from the consistent aggregation of the individual optimal decisions.

We let $\mathbb{E}^m(\bar{\pi}, \bar{w}^\ell)$ denote the set of sequences $\bar{\pi}^a$ that are consistent with an assimilation equilibrium given a pair of strategies $(\bar{\pi}, \bar{w}^\ell)$ for the leader and the majority group. An application of Proposition A.3 in the Appendix guarantees that such sequences exist. The share of individuals in the minority group who decide to assimilate is defined from

$$\phi_t^a(\theta^t, \bar{\pi}^a, \bar{w}^\ell, \bar{\pi}) = \int_{\mathcal{K}^a} \hat{g}_t^a(K^a, \theta^t, \bar{\pi}^a, \bar{w}^\ell, \bar{\pi})[\pi_t^a(\theta^{t-1})](\mathrm{d}K^a), \tag{11}$$

for all $\theta^t \in \Theta^t$ and $t > 0$. We then have

Proposition 2. *Let $\phi_{\vee,t}^a$ be defined from (11) for $(\bar{\pi}, \bar{w}^\ell)$ equal to the greatest discrimination equilibrium with a leader $(\bar{\pi}_\vee, \bar{w}_\vee^\ell)$ in E^M and $\bar{\pi}^a$ equal to the greatest assimilation equilibrium $\bar{\pi}_\vee^a$ in $\mathbb{E}^m(\bar{\pi}_\vee, \bar{w}_\vee^\ell)$ Let $d_t^\ell \in \{0,1\}$ denote the observed leader's action at time $t > 0$ If u^a exhibits increasing differences in (a, θ), (K^a, θ), (a, p), and (K^a, p), then*

$$\mathbb{E}[\phi_{\vee,t+\tau}^a | \theta^t, d_t^\ell = 1] \geq \mathbb{E}[\phi_{\vee,t+\tau}^a | \theta^t, d_t^\ell = 0]$$

for all $t > 0$, $\tau > 0$, and all $\theta^t \in \Theta^t$, where the expectation is taken with respect to density of $\theta^{t+\tau}$ conditional on θ^t and d_t^ℓ. If u^a exhibits decreasing differences in (a, θ), (K^a, θ), (a, p), and (K^a, p), then the opposite inequality holds. These inequalities also hold for the least assimilation equilibrium in $\mathbb{E}^m(\bar{\pi}_\vee, \bar{w}_\vee^\ell)$. Further, they hold for the greatest and least assimilation equilibria at $(\bar{\pi}_\wedge, \bar{w}_\wedge^\ell) \in E^M(\bar{s})$, the least discrimination equilibrium with a leader.

When the leader discriminates, the minority group anticipates that the leader's discrimination capital is likely to increase in the future, and that the collective response of the majority group will also make the environment more discriminatory. Then, the assimilation response in the aftermath of a leader's discriminatory action depends on whether the incentives to assimilate are stronger or weaker in a more discriminatory society. However, the result only holds in expectation because the actual sequence of leader's identities that will be observed is uncertain.

2.8 Comparative Statics for the Majority Group

In this paper, we interpret the outbreak of COVID-19 in the United States as an unexpected increase in the sequence $\bar{s} \in \mathcal{S}^{\infty}$. We are therefore interested in studying the effect of this shock to $\bar{s}$ on the equilibrium outcomes of the model, taking into account the interaction between the majority group and the leader.

Proposition 3. *Let $\bar{s}^2$, $\bar{s}^1$ be two sequences in $\mathcal{S}^{\infty}$ with $\bar{s}^2 \succeq \bar{s}^1$, and let $(\bar{\pi}^2_{\vee}, \bar{w}^{\ell,2}_{\vee}) \in E^M(\bar{s}^2)$ and $(\bar{\pi}^1_{\vee}, \bar{w}^{\ell,1}_{\vee}) \in E^M(\bar{s}^1)$ be the corresponding greatest discrimination equilibria with a leader. Then,*

$$\mathbb{E}[\phi^d_{t+\tau}(\theta^{t+\tau}, \bar{\pi}^2_{\vee}, \bar{w}^{\ell}_{\vee}, \bar{s}^2)|\theta^t] \geq \mathbb{E}[\phi^d_{t+\tau}(\theta^{t+\tau}, \bar{\pi}^1_{\vee}, \bar{w}^{\ell,1}_{\vee}, \bar{s}^1)|\theta^t]$$

for all $t > 0$, $\tau > 0$, and $\theta^t \in \Theta^t$, where the expectations are taken with respect to the density of $\theta^{t+\tau}$ conditional on θ^t. The same inequality holds for the least equilibria in $E^M(\bar{s}^2)$ and $E^M(\bar{s}^1)$.

If the equilibrium is unique, this proposition says that the share of individuals in the majority group who decide to discriminate against the minority increases after the exogenous circumstances become more favorable to discrimination. The effect arises from the combination of several channels. First, the individual incentives to discriminate become stronger. Second, in general equilibrium, this effect is amplified by the aggregate response of the majority group, because the distribution of discrimination capital shifts along all future histories of leader's identities. Third, the leader also has a stronger incentive to discriminate. Because the majority group anticipates that the leader will be more likely to discriminate in the future, the impact of a shock to $\bar{s}$ is further amplified by the interaction with the leader. Again, individual agents are unable to perfectly forecast the behavior of the leader, therefore the result holds in expectation.

2.9 Comparative Statics for the Minority Group

Next, we need to understand how the minority group reacts to changes to the discriminatory environment. In principle, it is not obvious whether the incentive to assimilate should be stronger or weaker in a more discriminatory environment. The empirical evidence on the impact of discrimination on assimilation is mixed: some studies show that increased discrimination can favor the assimilation of minorities (Fouka, 2019; Jaschke et al., 2022), others find that discrimination or forced assimilation policies can have the opposite effect (see Gould and Klor (2016b) and Fouka (2020b), respectively). In light of this observation, we do not take a stand on how the discriminatory environment enters the utility function of the minority group. Instead, we characterize the conditions under which, within our framework, a more discriminatory environment has a positive or negative effect on the assimilation of minorities.

Proposition 4. *Let $\bar{s}^1$, $\bar{s}^2$ be two sequences in $\mathcal{S}^\infty$ with $\bar{s}^2 \succeq \bar{s}^1$, let $(\bar{\pi}^1_\vee, \bar{w}^{\ell,1}_\vee) \in E^M(\bar{s}^1)$ and $(\bar{\pi}^2_\vee, \bar{w}^{\ell,2}_\vee) \in E^M(\bar{s}^2)$ be the corresponding greatest discrimination equilibria with a leader, and let $\bar{\pi}^{a,1}_\vee$ and $\bar{\pi}^{a,2}_\vee$ be the greatest assimilation equilibria in $\mathbb{E}^m(\bar{\pi}^1_\vee, \bar{w}^{\ell,1}_\vee)$ and $\mathbb{E}^m(\bar{\pi}^2_\vee, \bar{w}^{\ell,2}_\vee)$. If u^a has increasing differences in (a, θ), (K^a, θ), (a, p) and (K^a, p), then*

$$\mathbb{E}[\phi^a_{t+\tau}(\theta^{t+\tau}, \bar{\pi}^{a,2}, \bar{\pi}^2)] \geq \mathbb{E}[\phi^a_{t+\tau}(\theta^{t+\tau}, \bar{\pi}^{a,1}, \bar{\pi}^1)]$$

for all $t > 0$, $\tau > 0$, and $\theta^t \in \Theta^t$, where the expectations are taken with respect to the density of $\theta^{t+\tau}$ conditional on θ^t. If u^a has decreasing differences in (a, θ), (K^a, θ), (a, p) and (K^a, p), then the inequality holds with the opposite sign. The inequalities also hold for the least assimilation equilibria in $\mathbb{E}^m(\bar{\pi}^1_\vee, \bar{w}^{\ell,1}_\vee)$ and $\mathbb{E}^m(\bar{\pi}^2_\vee, \bar{w}^{\ell,2}_\vee)$. Further, they hold for the greatest and least assimilation equilibria at $(\bar{\pi}^1_\wedge, \bar{w}^{\ell,1}_\wedge) \in E^M(\bar{s}^1)$ and $(\bar{\pi}^2_\wedge, \bar{w}^{\ell,2}_\wedge) \in E^M(\bar{s}^2)$, the least discrimination equilibria with a leader.

In this proposition, we are studying the equilibrium assimilation behavior under two different discrimination equilibria with a leader. If the equilibrium is unique, Proposition 3 tells us that the sequence of distributions of discrimination capital will shift upwards and the leader will be more prone to discrimination along any future history of leader's identities. The dynamic assimilation response of the minority group depends on whether the incentive to assimilate is stronger or weaker in a more discriminatory society.

2.10 Discussion

We conclude this section by discussing the main mechanisms at work in the model. To ease the discussion, suppose that the equilibrium of the model is unique. Then, Proposition 3 says that the share of discriminating individuals increases after an expected increase to $\bar{s}$. This is due to four channels. First, there is a direct effect on the current payoff for discriminating. Second, since individuals are forward-looking, there are dynamic gains from discriminating, due to the fact that discrimination reinforces the individual's future identity. Third, individuals correctly anticipate that the social environment is going to be more discriminatory in the future. Fourth, the majority group also perceives that the leader is more likely to discriminate in the future—this incentive starts to play in even before the leader actually decides to discriminate.

The leader, for her part, is more likely to discriminate for two reasons: first, like ordinary individuals of the majority group, her direct payoff for discriminating is higher when $\bar{s}$ is higher; second, and more important, she realizes that the majority group is more prone to discrimination when circumstances $\bar{s}$ are favorable to it. This result arises from the strategic interactions between the leader and the majority group, as well as from the cross-complementarities assumed between the discriminatory environment, the leader's behavior, and the majority group members' individual discrimination choices.

Even a *one-period* shock to $\bar{s}$ has prolonged effects in this model, because the identities of those who discriminate change in the next period. This will also be reflected in the discriminatory environment, which contributes to protracting the consequences of the initial shock. Similarly, the effects of the leader's choice to discriminate also last more than one period, because the leader herself accumulates capital and because of the response her actions trigger in the majority group.

Finally, when $\bar{s}$ unexpectedly increases, the minority group immediately realizes two things: one, the discriminatory environment is going to be higher at all future dates, and two, the leader is also more likely to discriminate in the future. According to Proposition 4, the assimilation reaction of the minority group then depends on whether the incentives to assimilate are stronger or weaker in a discriminatory society. This is ultimately an empirical question.

3 Background and Data

In this section, we discuss the background and data used for the empirical analysis. In Section 3.1, we provide details on the U.S. context during the period from January 6, 2020 to August 27, 2020, when the U.S. population experienced an exogenous change to the returns to discrimination actions due to (1) the outbreak of a deadly virus that originated in China, and (2) a discriminatory action by the U.S. President (the most prominent political leader at that time) toward the Chinese community.[11] Then, in Section 3.2, we discuss how we constructed the sample and main dependent variables to study discrimination and assimilation behaviors following the two shocks.

3.1 Background

The 2019 novel coronavirus (SARS-CoV-2) was first identified in Wuhan, Hubei, China, where a major local outbreak suddenly became a global public health emergency. The first COVID-19 case in the United States was reported on January 20, 2020, in the state of Washington (Holshue et al., 2020). Italy was the first severely hit country in Western Europe, starting from late February. Following the start of the Italian national lockdown, the recording of cases in more than 100 countries, and the declaration of state of emergency in Rhode Island and Ohio, on March 9, several restrictive measures were adopted in the U.S. These included isolation and quarantine for several suspected and confirmed cases, cancellation of public events, and suspension of in-person classes at universities. On March 11, the World Health Organization declared the outbreak of a pandemic.

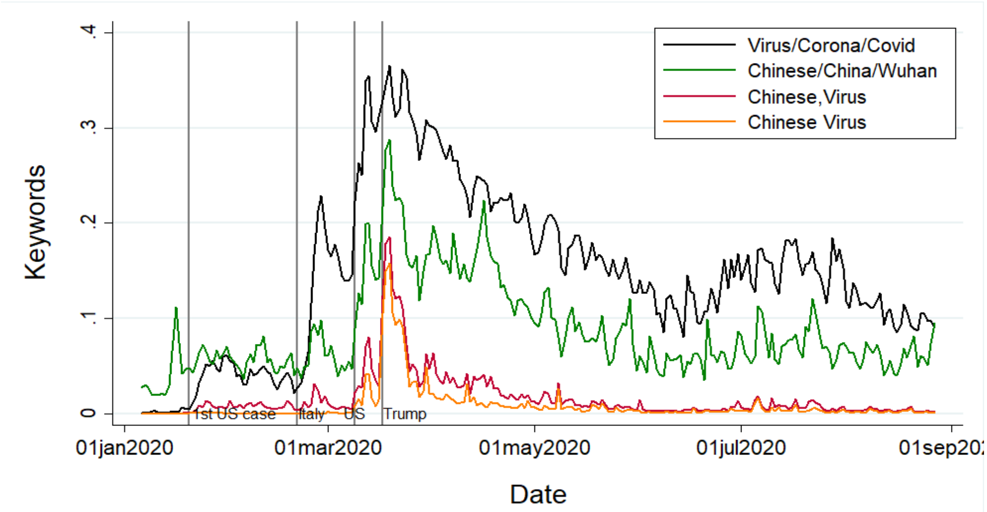
On the evening of March 16, U.S. President Donald Trump referred to the novel coronavirus as the "Chinese Virus" in a tweet regarding economic support for U.S. industries affected by the pandemic. One day later, National Public Radio announced that all U.S. states had reported COVID cases (NPR, 2020), and Trump reiterated his views on the "Chinese Virus," despite media accusations of racism (see Appendix D.1).

In line with these accounts, based on data from Twitter (see Section 3.2 for details on the data and the sample), the upper line in Figure 1 displays the evolution of the daily share of Twitter users tweeting keywords such as "virus," "corona," and "covid," and highlights the increasing salience of COVID-related topics in the overall tweeted text after the major events discussed above. The vertical lines labeled "1st US case", "Italy", "US", and "Trump" signal the first U.S. case on January 20, the outbreak in Italy on February 21, the implementation of restrictive measures in the United States from March 9, and the former president's series of discriminatory tweets from

[11] Based on the functionalities of the Twitter API accessed in February 2020, we could retrieve only the previous 3200 tweets of each user's timeline. To ensure homogeneous coverage for all users, we chose January 6, 2020 as the starting date for our sample. We end our period of analysis on August 27, 2020, when the number of U.S. COVID-19 hospitalizations reached the lowest point since the beginning of the pandemic.

March 17, respectively. Notably, in correspondence of both March 9 and March 17, we observe a sharp increase in the share of tweeted text containing keywords related to China (green line), and more specifically containing keywords related to both China and the virus (red and orange lines, the latter restricting the pattern to the use of the specific expression "Chinese virus"). Our analysis will precisely focus on March 9 and 17 as thresholds at which, according to the theory, we may find discontinuity both in discriminatory behavior and in the reaction of the minority discriminated against.

Figure 1: Daily Share of Users (all groups) Who Tweeted Selected Keywords



Notes: This figure represents the share of users who tweeted a given keyword or set of keywords on each date from January 6 to August 27, 2020. The gray vertical lines labeled "1st US case" (January 20), "Italy" (February 21), "US" (March 9) and "Trump" (March 17) signal, respectively, the first U.S. case, the outbreak in Italy, the implementation of restrictive measures in the United States, and the series of discriminatory tweets by Trump. The black line represents the daily share of users who tweeted *virus, corona,* or *covid.* The green line represents the daily share of users who tweeted *chinese, china,* or *wuhan.* The red line represents the daily share of users who tweeted *chinese* and *virus* in the same tweet, in any position or order. The yellow line represents the daily share of users who tweeted *chinese virus.*

3.2 Data

Our analysis is based on a rich dataset on social media activity in the United States, including discriminatory attitudes of the White group and related reactions of the Chinese group. In particular, within Twitter, we identify 8,130 White users and 942 Chinese users, and we follow their activity over time. For all users in the sample, we observe the universe of tweets from January 6, 2020, to August 27, 2020, a total of 5,576,061 tweets.

We select users based on three specific criteria: (1) the self-descriptions reported in their profiles must include keywords that signal their belonging to either the (non-Hispanic) White community or the Chinese community; (2) they have to be located in the United States during the study period (based on their self-reported location and that of their friends); (3) they have to be sufficiently active in the Twitter community and likely to be engaged in social, political, and cultural debates (see Appendix D.2 for details on how we constructed the dataset).

As in most studies based on social media users, the sample construction may involve some selection, as users tend to cluster (and post content) based on the type of information they are exposed to (see, for instance, the discussion in Schmidt et al., 2017; Sunstein, 2018; Müller and Schwarz, 2021). However, in our case selection concerns are mitigated by three observations. First, as self-reported in their Twitter bios, the users in our sample are demonstrably heterogeneous: full-time mothers, entrepreneurs, actors, politicians, and activists, among others. Second, White and Chinese users are also very heterogeneous with respect to, respectively, their preshock share of generalized abusive language (not directly against the Chinese community) and their share of U.S. friends (see Appendix Figure D5). Third, our focus on specific subsamples of users does not affect our baseline results (see Section 5). Altogether these observations suggest that our results are unlikely to be confounded by a specific subset of users, helping mitigate any selection concerns.

In addition, unlike survey or experimental data, social media data have the clear advantages of providing large samples and high-frequency measures of both discrimination and assimilation.

Our goal is to extract information on the discriminatory and assimilation intentions of the tweets posted by the users in our sample. To this aim, we adopt three text-analysis approaches: dictionary-based, supervised machine learning, and unsupervised machine learning. In the first two approaches, a textual unit of analysis, i.e., a "document," is a single tweet. In the third approach, we define a document to contain the entire corpus of text tweeted by a single user in a day; thus, if a user created multiple tweets in a day, we paste them together. The reason is that commonly used unsupervised methods, such as the Latent Dirichlet Allocation, tend to underperform with short texts.[12] We now briefly describe each of the three methods. For more details, see Appendix sections D.3, D.4, and D.5.

[12] See Hong and Davison (2010). Yan et al. (2013) propose an alternative algorithm, the Biterm Topic Model, to address this issue. Our results are robust to the use of Biterm analysis and are available upon request.

We start by performing dictionary-based exercises; that is, we search specific keywords inside the textual units of analysis. For instance, we look for anti-Chinese slurs (e.g., "ching chong") to detect anti-Chinese discriminatory tweets. Though simple and intuitive, this method bears the limitations that the choice of keywords is subject to human bias and that the context in which the keyword occurs is not taken into account.

Supervised methods, on the other hand, are based on a training dataset where the features of interest are observed for a certain number of (textual, in our case) examples. This information can then be leveraged to predict the same features in the main dataset of the analysis (Hastie et al., 2009). As far as discrimination attitudes are concerned, we exploit the annotated dataset by Founta et al. (2018), which has been adopted by the recent growing literature on hate-speech detection, particularly hate-speech detection on Twitter.[13] Here, tweets are labeled as "hateful," "abusive," "normal," or "spam." Then, we classify our own sample of tweets using the Support Vector Machines (SVM) algorithm, which is widely used in text-analysis applications.

In contrast, to the best of our knowledge, no comparable dataset exists for assimilation-related content. We take a different route, following the methodology of Ash et al. (2021), which is based on "word embeddings." Word embeddings are representations of words in an $\mathcal{R}^K$ vector space, such that each dimension corresponds to an aspect of meaning. Similar words will be proximate to each other, and, generally, relationships between words will follow an internally consistent metric. Documents can also be vectorized as (standardized) sums of the vectorized words that are part of the document. We embed our sample of tweets by Chinese users in a 150-dimension vector space, and we compute their distance to the "assimilation dimension" of that space. To locate this dimension, we vectorize a sample of sentences containing clear assimilation content, mostly drawn from sociological studies on assimilation, such as Kibria (2000), reporting interviews with second-generation Asian American.

Finally, for the unsupervised machine-learning method, we use the Latent Dirichlet Allocation (LDA) algorithm, developed by Blei et al. (2003) (see also Hansen et al. (2018) for an early application in economics), to identify the latent topics in the corpus of documents for a given group (e.g., White American users), and to derive the topic composition of each document.[14]

[13]Alternative datasets are provided by Waseem and Hovy (2016), Davidson et al. (2017), Vidgen et al. (2020) and Ziems et al. (2020).

[14]Three parameters have to be set externally: the number of topics (k), plus the two hyperparameters (α and δ)

We preprocess the raw text in two ways. First, we convert the text to lowercase and remove URLs, mentions, punctuation, and numbers (except when tagging U.S. congressional bills), plus a number of minor adjustments, which we detail in Appendix D.2. Second, when we use the text as an input in the supervised and unsupervised methods, we subject it to three additional steps: (1) we remove stopwords; (2) we replace words with their stems (using the Porter stemmer); and (3) we apply the tf-idf (term frequency-inverse document frequency) filter, in line with Hansen et al. (2018). Also, before preprocessing, all text in Chinese (in the Chinese users' tweets) is translated to English via the DeepL API.

3.2.1 Discrimination

We now detail our measures of discrimination and how they are constructed. First, we perform a dictionary-based exercise, selecting a set of 18 keywords that express anti-Chinese racial slurs.[15] We compute a dummy variable labeled *Anti-Chinese Slurs* taking the value one if, on a given day t, the user tweeted at least one of these keywords. This represents our first discrimination outcome.

Second, we perform a supervised-machine-learning exercise that relies on the dataset of 99,996 annotated tweets presented in Founta et al. (2018) as the training dataset. These tweets were collected via the Twitter Stream API from March 30 to April 9, 2017, then classified into four categories: normal, abusive, hateful, and spam, via the *CrowdFlower* crowdsourcing platform. After deleting the spam tweets, we end up with 27,150 (32%) abusive tweets, 4,965 (6%) hateful tweets, and 53,851 (62%) normal tweets.

With this labeled dataset, we perform the following classification exercise: we group hateful and abusive tweets in the same class, then we predict hateful and/or abusive *versus* normal speech. In this exercise, we split the sample in a training set (including 70% of the tweets) and a test (including the remaining 30% of the tweets). We use the training set to train a Support Vector Machine model (see Hastie et al. (2009) for a reference). To optimize the SVC cost parameter, we use a fivefold cross-validation procedure.[16] Then we use the test set to evaluate the performance of the model.

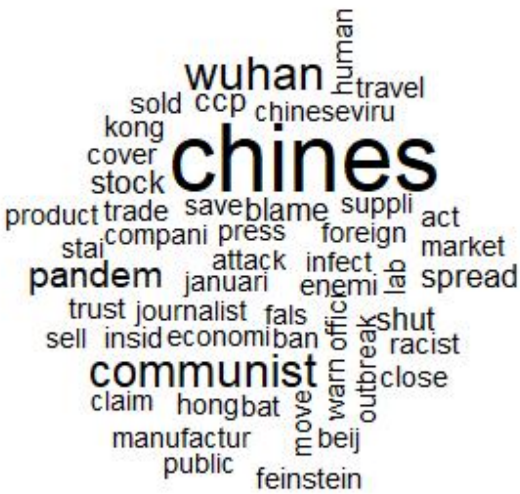
for the prior Dirichlet distributions. In the following exercises, we set $k = 60$ for the White group and $k = 40$ for the Chinese group, and, following Griffiths and Steyvers (2004) and Hansen et al. (2018), $\alpha = k/50$ and $\delta = 0.1$.

[15] The 18 keywords are *chink, chingchong, chinesebat, chinesestudentban, chinesespy, chonk, churka, cokin, coolie, dink, flango, gook, kungfuflu, niakoue, slanteye, slopehead, tingtong,* and *yokel.*

[16] Essentially, this procedure performs a grid search over different values of the tuning parameter and selects the one that achieves the best performance. To measure performance, the algorithm splits the training dataset in n (five, in our case) subsets (folds). Then it trains the model on a dataset comprising $n - 1$ folds and test its performance

On top of word frequencies, we use the number of letters and the number of punctuation marks as additional predictive features.[17] The model achieves an accuracy (share of correctly classified observations) of 92.1%.[18] We use this model to classify the tweets in our sample and compute the dummy variable *Abusive Language* taking the value one if on a given day t a user posted one or more hateful or abusive tweets that contained the keyword "Chinese." This is our second discrimination outcome.

Figure 2: The "Discrimination" Topic Wordcloud



Notes: Wordcloud for the "Discrimination" topic based on the LDA on the overall text tweeted by users of the White group from January 6 to August 27, 2020. Larger words are more recurrent. See Appendix D.5.1 for details.

Finally, we run the LDA algorithm on the corpus of daily tweeted documents of the White users. In our baseline exercise, we use 60 topics (we also find the main topic of interest using more or fewer topics). Figure D3 in Appendix D.5 shows the most representative words for each topic. For our empirical analysis, we focus on one particular topic of interest, which we naturally labeled "Discrimination" (topic 19), as the documents with highest weights across time were clearly discriminatory against the Chinese community (see examples in Appendix D.5.1). The most rele-

over the excluded fold; it repeats this step n times, such that a different fold is excluded in each iteration. At the end, it computes an average measure of performance. We use the following grid: $0.1, 1, 10, 100, 10000$, and the chosen parameter is 0.1.

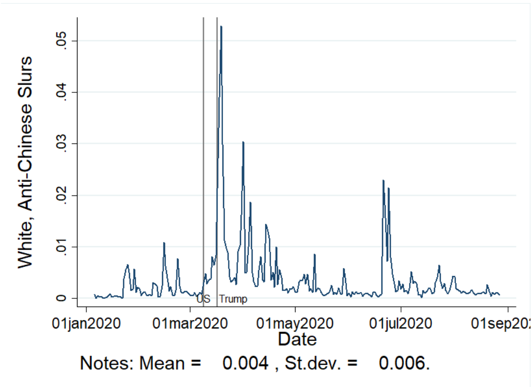
[17] In the Founta et al. (2018) dataset, the average number of letters is 94.3 in normal tweets, versus 85.4 in hateful tweets and 70.4 in abusive tweets; the average number of punctuation marks is 2.83 in normal tweets, versus 2.35 in hateful tweets and 2.13 in abusive tweets.

[18] We report the confusion matrix in Appendix D.3.

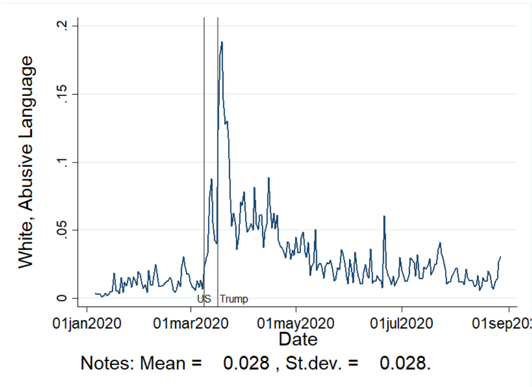
vant keywords for this topic are *Chinese, Wuhan, communist, pandemics, spread, stock, ccp, shut, blame,* and *lab.* Figure 2 shows the wordcloud for this topic. Our third discrimination outcome is the variable *Discrimination Topic*, which, for each day t, computes the average of the share of text pertaining to this topic within the entire sample of users tweeting in that given day (therefore, the number of observations corresponds to the number of days in the sample).

Figure 3 reports the evolution of the daily share of White users who tweeted anti-Chinese slurs, and expressions with abusive language, as well as the average of the daily share of text associated with the topic discriminating the Chinese community. At the bottom of each graph, we report the mean and standard deviation of the respective variable. Consistently across all outcome variables, the figure shows a peak in discriminatory language on March 9, 2020, and a subsequent sizeable spike on March 17, 2020.

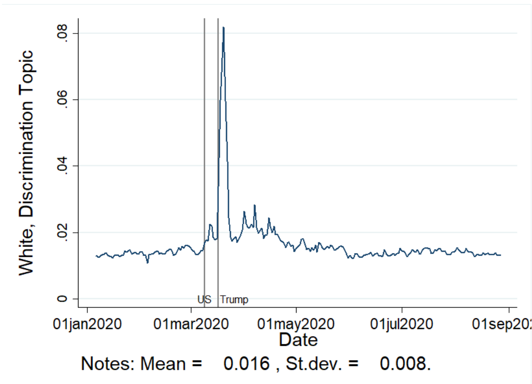
Figure 3: Discrimination Keywords, Expressions, and Topics for the White Users



(A) Anti-Chinese Slurs



(B) Abusive Language



(C) Discrimination Topic

Notes: Daily share of White users tweeting using Anti-Chinese slurs (Panel A) , abusive language against the Chinese (Panel B), and daily average of the share of text attributed to the selected topic within the sample of White American users tweeting in a given day (Panel C). All panels refer to the period from January 6 to August 27, 2020. Mean and standard deviation of the variables in the period of interest are reported below each graph.

3.2.2 Assimilation

We construct our proxies of assimilation attitudes by building on the sociology literature and, specifically, on the process of the identity formation of second-generation Asian Americans. In particular, Kibria (2000) examines the dynamics of race and ethnic identity for the Chinese and Korean ethnic groups in U.S. society and emphasizes two key assimilation behaviors of the second generation: (1) *establishing Americanness* and (2) *distancing from the reference ethnic group*. We construct two measures of assimilation, based on the first type of behavior, and a third proxy, based on the second type of behavior. Starting with the first assimilation behavior, Kibria (2000, p. 86) suggests that *establishing Americanness* by stating the ethnic identity (e.g, Chinese or Asian) followed by the nationality (i.e., American), or by expressing adherence to mainstream U.S. culture, norms, and practices, are key strategies for establishing one's identity as "American" and asserting "the nature of one's ties and relationship to US society and culture."[19] For the first measure of this type of behavior, we rely on the dictionary-based methodology and compute a dummy variable labeled *Chinese/Asian American*, that takes value one if, on a given day t, the user posted one or more tweets with the phrase "Chinese American" or "Asian American."

Then, to build our second measure of assimilation, we rely on a supervised-machine-learning approach to detect expressions pointing to a sense of belonging to the American community. In particular, we train a Word2vec algorithm on the corpus of Chinese tweets, and we use it to compute the cosine similarity between each tweet and a set of (vectorized) assimilation sentences (51 sentences, from five distinct sources, with a clear assimilation content—see Table D2 in Appendix D.4 for full text and sources). In these sentences, first- and second-generation Asian Americans express how they consider themselves as being of American nationality and the extent to which they have embraced the American culture, going beyond their ethnicity of origin. For instance, in one of the sentences, the interviewee declares that she would give herself the American label because that's where she has spent most of her life, and because the state recognizes her as a citizen. She also asserts that she considers herself culturally American. Another interviewee reported feeling

[19] "The notion of Asians as 'essential foreigners' has played an important role in organizing and legitimating hostility toward and discrimination against Asians in the United States [since] the 1882 Chinese Exclusion Act, which banned Chinese immigration to the United States. Observers also note how hate crimes against Asians today often focus on their presumed 'foreignness.' [...] Thus the second generation felt compelled [...] to downplay their distinctive ethnic backgrounds in order to establish themselves as 'American'." (Kibria, 2000, p. 86).

America as her familiar world and China as something she had to learn. For each tweet, we define a unique measure of similarity to assimilation content as the maximum of the sentence-specific similarity scores and classify the top 1% of tweets, in terms of similarity to assimilation sentences, as tweets with assimilation content.[20] Since the share of tweets with either "chinese american" or "asian american" is around 0.1% in our dataset, the 1% cutoff, albeit arbitrary, seems a reasonable benchmark. Therefore, we compute the dummy variable *Assimilation* as equal to one if, on a given day t, a user posted one or more tweets that are predicted to carry assimilation content according to this definition. This is our second measure of assimilation.

Kibria (2000, p. 84) also emphasizes that *distancing from the reference ethnic group* has traditionally been a way to "cope with racialized hostility." Disidentification, ("the use of various clues and signals to distance oneself from the perceived 'problem group'" (Goffman, 2009)) often helped second-generation Chinese and Korean Americans to deflect hostility when accused by strangers in public spaces for the problems caused by various Asian groups. Similarly, in our case, distancing oneself from some distinctive features of China could be a strategy for facing anti-Chinese/Asian sentiments. Given that the CCP is a defining characteristic of China, and that the Party was particularly involved in managing the pandemic, distancing from the CCP seems a natural way to cope with discrimination in the COVID-19 context. To measure this, we use the LDA algorithm to model the daily tweets of the Chinese users into 40 topics.[21] Figure D4 in Appendix D.5 shows the set of words that mostly represent each topic. For our empirical analysis, we focus on a particular topic of interest, which can robustly be found also when we increase or decrease the number of topics to be retrieved by the algorithm. We naturally labeled topic 13 as *Blame CCP*—for which the relevant keywords are *ccp, communist, party, ccpvirus, truth, hsk, american, govern, pandem,* and *america*—and we summarize text that is largely devoted to accusing the CCP for the spread of the virus in China, in the United States, and around the globe.

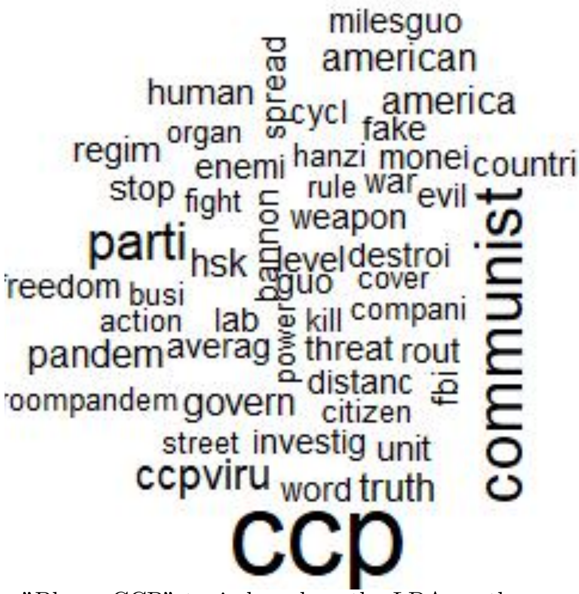
Figure 4 displays the resulting wordcloud. Examples of text classified into this topic are reported in Appendix D.5.2. Our third assimilation outcome is the variable *Blame CCP Topic*, which, for each day t computes the average of the share of text pertaining to this topic within the entire

[20] Ash et al. (2021) use the cosine similarity directly as a dependent variable; here, we categorize this variable in order to be consistent with the rest of our analysis.

[21] Because our sample of Chinese users is smaller than our sample of White users, we employ a lower number of topics for the unsupervised-machine-learning exercise on the Chinese group.

sample of users tweeting on that day (the number of observations corresponds to the number of days in the sample).

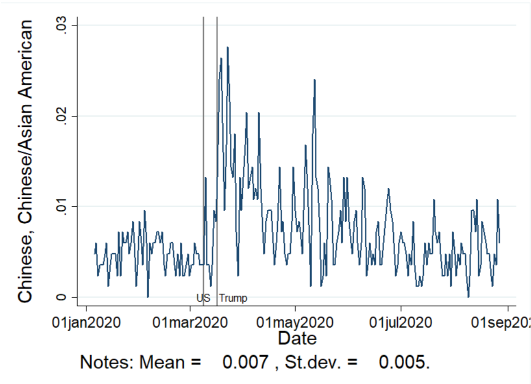
Figure 4: The "Blame CCP" Topic Wordcloud



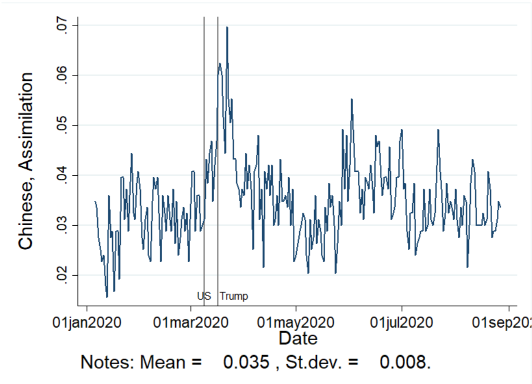
Notes: Wordcloud of the "Blame CCP" topic based on the LDA on the overall text tweeted by users of the Chinese group from January 6 to August 27, 2020. Larger words are more recurrent. See Appendix D.5.2 for details.

Figure 5 reports the evolution of the daily share of Chinese users who tweeted the selected keywords (Chinese/Asian American) and content (assimilation-like), as well as the average of the daily share of text associated with blaming the CCP within the sample of Chinese users. At the bottom of each graph, we report the mean and standard deviation of the respective variable in the sample. Consistently across all outcome variables, the figure shows a rise in assimilation language after March 9, 2020, and March 17, 2020, the latter increase being substantially more pronounced.

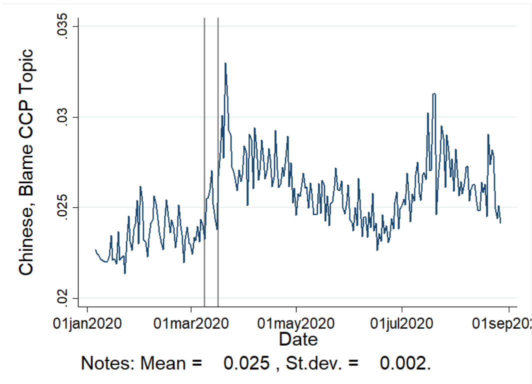
Figure 5: Discrimination Keywords, Expressions, and Topics for the Chinese users



(A) Chinese/Asian American



(B) Assimilation



(C) Blame CCP Topic

Notes: Daily share of Chinese users tweeting selected keywords and expressions (Panels A and B), and daily average of the share of text attributed to the selected topic within the sample of Chinese users tweeting in a given day (Panel C). All panels refer to the period from January 6 to August 27, 2020. Mean and standard deviation of the variables in the period of interest are are reported below each graph.

4 Empirical Specification

We now present our empirical strategy to estimate the causal effect of both the health and leader's discrimination shocks on discriminatory and assimilation attitudes of the White and Chinese groups. In terms of the theory, the COVID 19 outbreak corresponds to an unexpected increase of the parameter $\bar{s}$ (see Propositions 3 and **??**), whereas Trump's Chinese-virus tweet corresponds to a leader's discriminatory action (see Lemmas **??** and **??**).

In particular, we adopt two different specifications. We start by considering the following linear trend break model:

$$\begin{aligned} Y_{it} = \beta_0 + \beta_1 t + \beta_2 FromMarch9 + \beta_3 FromMarch17 \\ +\beta_4 FromMarch9 \times [t - March9] \\ +\beta_5 FromMarch17 \times [t - March17] \\ +\phi X_{it} + \varepsilon_{it} \end{aligned}$$

where Y_{it} is the probability of tweeting a certain keyword or content for user i in day t, or the average share of text about a certain topic for the users of a given group in day t. The main coefficients of interest are β_2 and β_3, which measure the intercept changes in the relationship between the dependent variable and the time upon occurrence of the health and discriminatory shocks, respectively. We control for linear time trends, which switch on from March 9 and March 17. The set of controls X_{it} includes day-of-the-week and month-of-the-year dummies to account for possible day- and month-specific tweeting patterns. We cluster standard errors at the treatment level, that is by date.

Next, using a regression discontinuity design, we estimate the impact of the two shocks using two truncated samples of dates. When focusing on March 9, we consider all days before March 17; when focusing on March 17, we concentrate on all days after March 9. While this approach results in an effectively narrower band, it should ensure that the estimated effect of each shock is not "confounded" by the effect of the other shock. The estimating equations will be

$$Y_{it} = \beta_0 + \beta_1 FromMarch9 + f(t) + \phi X_{it} + \varepsilon_{it}$$

or

$$Y_{it} = \beta_0 + \beta_1 FromMarch17 + f(t) + \phi X_{it} + \varepsilon_{it}$$

where in both equations the coefficient of interest is β_1. The forcing variable is t, and $f(t)$ is a polynomial function in the forcing variable with different coefficients on each side of the cutoff date (March 9 and 17, respectively). As in the trend-break model, the set of controls X_{it} includes day-of-the-week and month-of-the-year dummies to account for possible day- and month-specific tweeting patterns. Note that our running variable, date, is somewhat imperfectly measured, since tweets are posted at specific hours and minutes of the day, resulting in a discrete rather than a continuous score. In this case, we follow Lee and Card (2008) and cluster the error term by date in the RD framework as well.[22]

5 Results

5.1 Increasing Discrimination by "White Americans"

Figure 6 illustrates our estimates of the impact of a health shock and a leader-induced shock on our measures of discriminatory attitudes, both for the trend-break model (graph a) and for the regression discontinuity strategy (graphs b and c). In Panel A, the dependent variable is a dummy equal to one if a user tweeted anti-Chinese slurs on a given date; in Panel B, the dependent variable is a dummy equal to one if, on a given date, a user posted a tweet containing both abusive language and the word *Chinese*; finally, in Panel C, the dependent variable is the daily average of the share of text related to the "Discrimination" topic, computed within the sample of users tweeting on a given day. Two main results emerge. First, in line with Proposition 3, discriminatory behavior, as proxied by our three measures, increases after March 9, following the COVID-19 outbreak in the United States. Second, as predicted by Proposition 1, discrimination against Chinese people also spikes after Trump's tweet on the "Chinese virus." Empirically, we find this effect to be even larger than the effect of the COVID-19 shock.

We report formal estimates in Table D5 in Appendix D.7.1 for the trend-break model and in Table 1 for the regression discontinuity approach.[23] In both tables, odd columns report unconditional estimates, while even columns report our preferred specifications controlling for dummies for days of the week and months of the year. In particular, column 4 of Table 1 shows that the probability of tweeting abusive language against Chinese people increases by more than 2 percentage points on March 9 (Panel A), and by almost 12 percentage points on March 17 (Panel B). Similarly, the average share of text related to anti-Chinese discrimination jumps when the health-

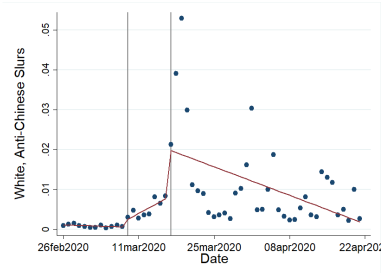
[22] Alternatively, Cattaneo et al. (2018) suggest using a local randomization approach. Results employing this method (available upon request) align with baseline results of this paper.

[23] RDD estimates are based on the truncated samples as described in Section 4. We estimate bandwidth based on the mean squared error (MSE) procedure allowing for different bandwidths on each side of the cutoff. Using the RDD optimal bandwidth we defined the sample for the trend-break model estimates.

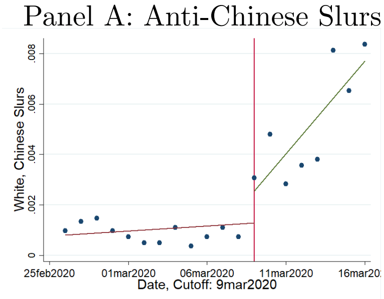
and leader-induced shocks take place (+0.4 and +5.0 percentage points, respectively), with a more precisely estimated and larger magnitude on March 17.

Figure 6: Tweeting “Anti-Chinese Slurs,” “Abusive Language,” and “Discrimination” Topics in Time

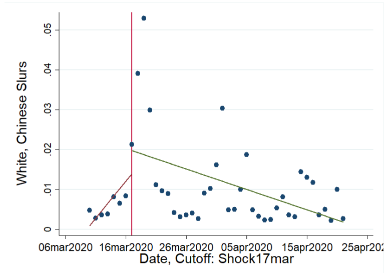
Panel A: Anti-Chinese Slurs



(a) Trend-Break Model

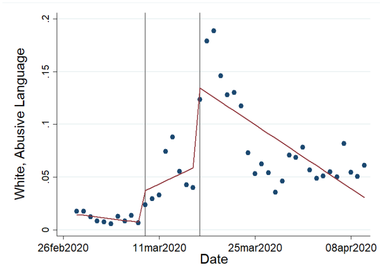


(b) RDD Cutoff March 9

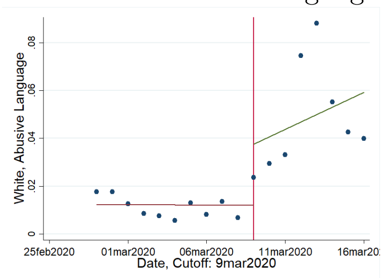


(c) RDD Cutoff March 17

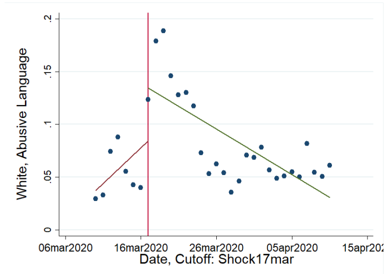
Panel B: Abusive Language



(a) Trend-Break Model

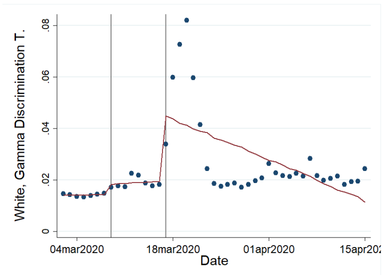


(b) RDD Cutoff March 9

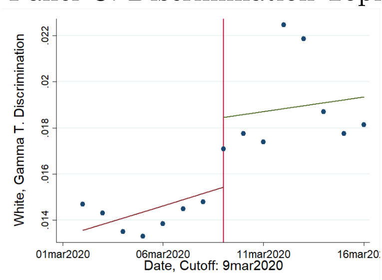


(c) RDD Cutoff March 17

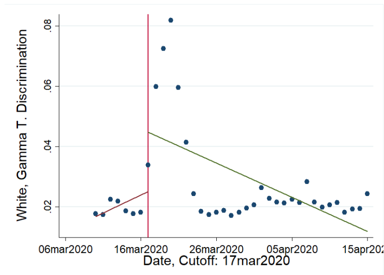
Panel C: Discrimination Topic



(a) Trend-Break Model



(b) RDD Cutoff March 9



(c) RDD Cutoff March 17

Notes: We consider the sample of ”White American” users. In Panel A, the dependent variable is a dummy taking the value 1 if the user tweeted anti-Chinese slurs; in Panel B, it is a dummy taking the value 1 if the user posted a tweet containing both abusive language and the keyword *Chinese*. Dots represent averages of the dependent variable (y-axis) in each day (x-axis), while continuous lines are unconditional linear fits on the panel of user-day observations. In Panel C, the dependent variable is the average of the share of text on the topic “Discrimination” against Chinese people computed within the sample of users who tweeted that day. Dots represent the values of the dependent variable (y-axis) in each day (x-axis), while continuous lines are unconditional linear fits. In each panel, graph (a) depicts trend-break-model estimates, graph (b) shows RDD results using March 9 as the cutoff, and graph (c) shows RDD results using March 17 as the cutoff. See Section 3 and Appendix D.2 for details on data construction and sources.

Table 1: Discrimination: RDD Estimates

Panel A: White Sample, Cutoff March 9, 2020						
Dep. Var.	Anti-Chinese Slurs		Abusive Language		Discrimination T.	
	(1)	(2)	(3)	(4)	(5)	(6)
RD_Estimate	0.0020	0.0019	0.0309	0.0225	0.0042	0.0040
	(0.0007)	(0.0005)	(0.0107)	(0.0076)	(0.0012)	(0.0012)
Robust P-value	0.0011	0.0071	0.3445	0.0081	0.0588	0.1152
Observations Left	97560	89430	81300	65040	7	6
Observations Right	65040	48780	65040	48780	8	8
Polynomial Order	1	1	1	1	1	1
Band. Method	msetwo	msetwo	msetwo	msetwo	msetwo	msetwo
Band. Left	12.556	11.776	10.675	8.154	7.293	6.717
Band. Right	7.000	5.087	7.000	5.953	7.000	7.000
Day and Month Dummies		✓		✓		✓

Panel B: White Sample, Cutoff March 17, 2020						
Dep. Var.	Anti-Chinese Slurs		Abusive Language		Discrimination T.	
	(1)	(2)	(3)	(4)	(5)	(6)
RD_Estimate	0.0111	0.0191	0.0781	0.1169	0.0260	0.0503
	(0.0048)	(0.0052)	(0.0193)	(0.0137)	(0.0073)	(0.0066)
Robust P-value	0.0395	0.0003	0.0000	0.0000	0.0001	0.0000
Observations Left	56910	56910	56910	56910	7	7
Observations Right	292680	186990	203250	146340	30	21
Polynomial Order	1	1	1	1	1	1
Band. Method	msetwo	msetwo	msetwo	msetwo	msetwo	msetwo
Band. Left	7.000	7.000	7.000	7.000	7.000	7.000
Band. Right	35.161	22.124	24.782	17.869	29.807	20.059
Day and Month Dummies		✓		✓		✓

Notes: We consider the sample of tweets of "White American" users before March 17 in Panel A and after March 9 in Panel B. In columns 1–2 of Panels A and B, the unit of observation is the user-day and the dependent variable is a dummy taking the value 1 if the user tweeted anti-Chinese slurs. In columns 3–4 of Panels A and B, the unit of observation is the user-day and the dependent variable is a dummy taking the value 1 if the user posted a tweet containing both abusive language and the keyword *Chinese*. In columns 5–6 of Panels A and B, the unit of observation is the day and the dependent variable is the average of the share of text on the topic "Discrimination" against Chinese people computed within the sample of users who tweeted that day. Results are local polynomial estimates using March 9 as the cutoff in Panel A and March 17 as the cutoff in Panel B, odd specifications are unconditional, and even specifications control for dummies for days of the week and months of the year. Standard errors, clustered by date in columns 1–4 and robust in columns 5–6, are reported in parentheses. Statistical significance is computed based on the robust P value. Different bandwidths on each side of the cutoff are derived under the MSE procedure using a linear polynomial and a uniform kernel.

Robustness. A possible concern could be that our results are driven by specific sets of users included in the analysis. For instance, if all users in our sample were intensely involved in politics or were active on social issues, our results would be valid only for this specific sample of the population. To alleviate this concern, we modify the White sample in two main ways. First, we check that our results are robust to excluding politically engaged and activist users based on the keywords in their bio (i.e., "republican," "democratic," "liberal," "feminist," "dissident," or "human rights").[24] Table D6 shows that the findings hold, despite having slightly lower magnitudes. Second, in Table D7, we also show that coefficients are quite stable and statistically significant when focusing on the specific subset of users who include political or activist keywords in their bio.

Overall, in line with the predictions of the model, the findings in this section point to an increase in discriminatory attitudes of the majority group after both health- and leader-induced shocks. In the following section, we assess the reaction of the Chinese minority.

5.2 Reaction of the Chinese Minority

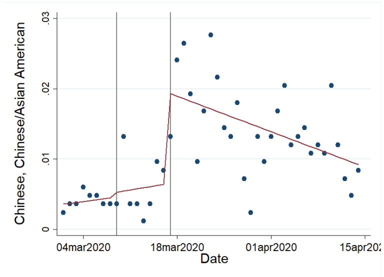
We now empirically investigate how Chinese people reacted to the worsening discriminatory environment. Based on Propositions 4 and 2, an increase in the returns to discriminate may have a positive or a negative influence on Chinese assimilation. In particular, Figure 7 depicts the results of the trend-break model (graph a) and regression discontinuity designs (graphs b and c) to estimate the effect of the COVID-19- and leader-induced discrimination on the assimilation attitudes of Chinese people.

Panel A of Figure 7 displays results using as dependent variable a dummy equal to one if the user tweeted using the keywords "Chinese American" or "Asian American"; Panel B uses a dummy equal to one if the user tweeted assimilation content; and Panel C displays results using the daily average of the share of text devoted to the topic "Blame CCP," computed within the sample of users tweeting in a given day. None of the dependent variables (except for "Blame CCP") show a significant jump when the COVID-19 shock occurs (March 9), while there is a positive and significant increase on March 17, following Trump's tweet on the "Chinese virus."

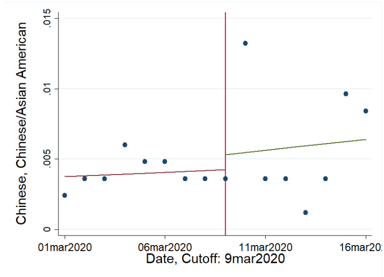
[24]See Appendix D.2 for the full list of keywords we use to identify this subsample of users.

Figure 7: Tweeting "Chinese/Asian American," "Assimilation," and "Blame CCP" topics in Time

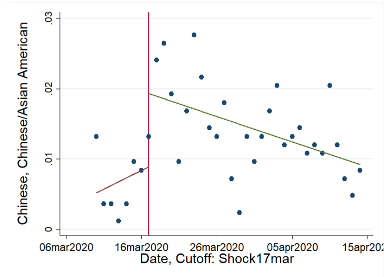
Panel A: Chinese/Asian American



(a) Trend-Break Model

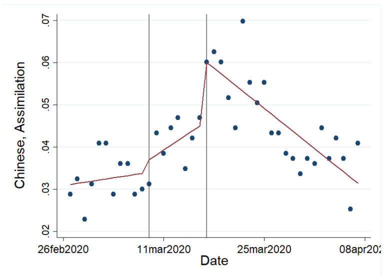


(b) RDD Cutoff March 9

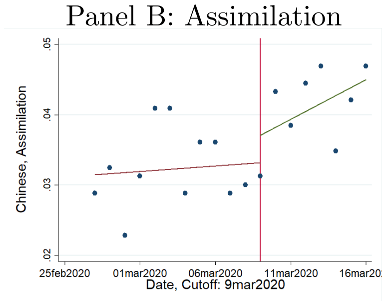


(c) RDD Cutoff March 17

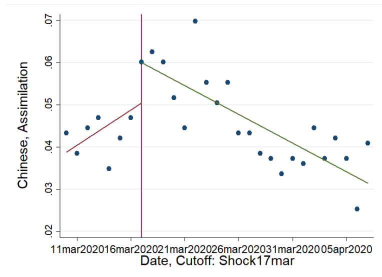
Panel B: Assimilation



(a) Trend-Break Model

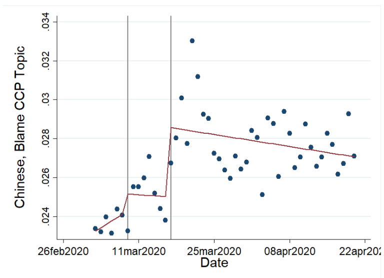


(b) RDD Cutoff March 9

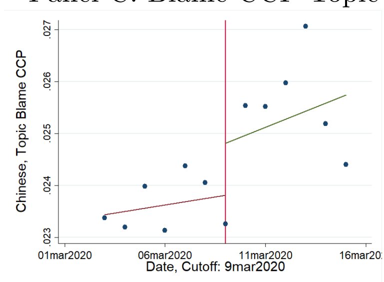


(c) RDD Cutoff March 17

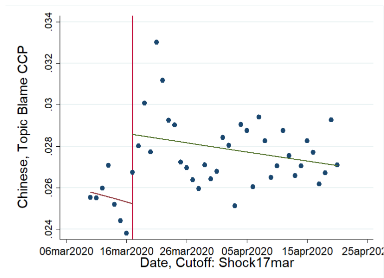
Panel C: Blame CCP Topic



(a) Trend-Break Model



(b) RDD Cutoff March 9



(c) RDD Cutoff March 17

Notes: Here, we consider the sample of Chinese users. In Panel A, the dependent variable is a dummy taking the value 1 if the user tweeted the keywords "Chinese American" or "Asian American"; in Panel B, it is a dummy taking the value 1 if the user tweeted assimilation content. Dots represent averages of the dependent variable (y-axis) in each day (x-axis), while continuous lines are unconditional linear fits on the panel of user-day observations. In panel C, the dependent variable is the average of the share of text on the topic "Blame CCP" computed within the sample of users who tweeted that day. Dots represent the values of the dependent variable (y-axis) in each day (x-axis), while continuous lines are unconditional linear fits. In each panel, graph (a) depicts trend-break-model estimates, graph (b) shows RDD results using March 9 as the cutoff, and graph (c) shows RDD results using March 17 as the cutoff. See Section 3 and Appendix D.2 for details on data construction and sources.

Table 2: Assimilation: RDD Estimates

Panel A: Chinese Sample, Cutoff March 9, 2020						
Dep. Var.	Chinese/Asian American		Assimilation		Blame CCP Topic	
	(1)	(2)	(3)	(4)	(5)	(6)
RD_Estimate	0.0007	-0.0029	0.0030	0.0018	0.0009	0.0002
	(0.0031)	(0.0021)	(0.0042)	(0.0022)	(0.0010)	(0.0005)
Robust P-value	0.4971	0.1385	0.8245	0.3497	0.1048	0.8052
Observations Left	6656	7488	9152	6656	6	6
Observations Right	6656	6656	6656	6656	8	8
Polynomial Order	1	1	1	1	1	1
Band. Method	msetwo	msetwo	msetwo	msetwo	msetwo	msetwo
Band. Left	8.625	9.001	11.969	8.635	6.615	6.813
Band. Right	7.000	7.000	7.000	7.000	7.000	7.000
Day and Month Dummies		✓		✓		✓

Panel B: Chinese Sample, Cutoff March, 17 2020						
Dep. Var.	Chinese/Asian American		Assimilation		Blame CCP Topic	
	(1)	(2)	(3)	(4)	(5)	(6)
RD_Estimate	0.0135	0.0181	0.0164	0.0213	0.0044	0.0049
	(0.0039)	(0.0034)	(0.0044)	(0.0033)	(0.0009)	(0.0008)
Robust P-value	0.5741	0.0111	0.0078	0.0000	0.0000	0.0000
Observations Left	5824	5824	5824	5824	7	7
Observations Right	24128	19136	18304	15808	35	30
Polynomial Order	1	1	1	1	1	1
Band. Method	msetwo	msetwo	msetwo	msetwo	msetwo	msetwo
Band. Left	7.000	7.000	7.000	7.000	7.000	7.000
Band. Right	28.004	22.449	21.579	18.287	34.389	29.942
Day and Month Dummies		✓		✓		✓

Notes: We consider the sample of tweets of Chinese users before March 17 in Panel A and after March 9 in Panel B. In columns 1–2 of Panels A and B, the unit of observation is the user-day and the dependent variable is a dummy taking the value 1 if the user tweeted the keywords "Chinese American" or "Asian American." In columns 3–4 of Panels A and B, the unit of observation is the user-day and the dependent variable is a dummy taking the value 1 if the user posted a tweet containing assimilation content. In columns 5–6 of Panels A and B, the unit of observation is the day and the dependent variable is the average of the share of text on the topic "Blame CCP" computed within the sample of users who tweeted that day. Our results are local polynomial estimates using March 9 as the cutoff in Panel A and March 17 as the cutoff in Panel B, odd specifications are unconditional, and even specifications control for dummies for days of the week and months of the year. Standard errors, clustered by date in columns 1–4 and robust in columns 5–6, are reported in parentheses. Statistical significance is computed based on the robust P value. Different bandwidths on each side of the cutoff are derived under the MSE procedure using a linear polynomial and a uniform kernel.

Table D8 in Appendix D.7.2 and Table 2 provide the formal estimates of the linear trend-break model and of regression discontinuity, respectively. Odd columns report unconditional estimates, while even columns report our preferred specifications controlling for dummies for days of the week and months of the year. None of the dependent variables show a significant jump when the COVID-19 shock occurs (March 9), while there is a positive and significant increase on March 17, following Trump's tweet on the "Chinese virus."[25] In particular, the political-discrimination shock is associated with a 1.81-percentage-point increase in the probability of tweeting "Chinese/Asian American" (see column 2), a 2.13-percentage-point increase in the probability of tweeting assimilation content (column 4), and a 0.49-percentage-point increase in the average share of text blaming the CCP (column 6).

Robustness. We present two sets of exercises showing the robustness of our results to using alternative proxies of *establishing Americanness* and alternative samples of Chinese users. In both cases—because our results in Figure 7 and Table 2 are mostly not significant on March 9—we focus on the March 17 shock.

First, in columns 1–2 of Table 3, we use alternative definitions of our first measure, the dummy tracking whether the user tweeted *Chinese* or *Asian* combined with *American.* A careful reading of the tweets reveals that when expressing particularly strong assimilation content, users also included the pronoun *we* together with *Chinese American* or *Asian American.* Therefore, in Column 1 of Table 3, our Chinese/Asian American dummy is equal to one only when the tweet also includes the keyword *we.* The result is robust to using this more restrictive measure of assimilation, although lower in magnitude with respect to column 2 of Table 2 (Panel B).

Next, one important concern is that our results could be driven by tweets using *Chinese American* or *Asian American* in the context of reported acts of discrimination against the Chinese and/or Asian communities. Thus, we set at zero our dummy if the tweet includes keywords such as *report, hate, spit, yell, incident, harass, anti-asian, blame, discriminate, affirmative action, hatred, attack,* or *scapegoat.* Column 2 shows that the result is robust.

We now focus on our second measure of *establishing Americanness*, based on the supervised-machine-learning exercise (columns 3–4). To rule out that tweets containing the keyword *feel* (and its variants) might report a general feeling rather than a sense of Americanness, in column 3 we set the dependent variable at zero if the text also included the keyword *feel* (and its variants). The

[25] This is consistent with the substantially larger increase in discrimination we find in correspondence of March 17 than in correspondence of March 9.

result is virtually unchanged with respect to column 4 of Table 2 (Panel B). Moreover, we changed the set of assimilation sentences at the basis of the word2vec algorithm to consider only sentences that include the keyword *American.* The result in column 4 is positive and significant, although slightly smaller in magnitude than in the previous exercises.

Table 3: Robustness Chinese/Asian American and Assimilation: RDD Estimates, Cutoff March 17

Dep. Var.	Ch./As.Amer.+We	Ch./As.Amer. No Report	Assim. No Feel	Assim. American
	(1)	(2)	(5)	(6)
RD_Estimate	0.0064	0.0103	0.0217	0.0145
	(0.0012)	(0.0024)	(0.0029)	(0.0038)
Robust P-value	0.0001	0.0276	0.0000	0.0042
Observations Left	5824	5824	5824	5824
Observations Right	24128	19136	14976	16640
Polynomial Order	1	1	1	1
Band. Method	msetwo	msetwo	msetwo	msetwo
Band. Left	7.000	7.000	7.000	7.000
Band. Right	28.579	22.693	17.306	19.192
Day and Month Dummies	✓	✓	✓	✓

Notes: We consider the sample of tweets of Chinese users after March 9. The unit of observation is the user-day. The dependent variable is a dummy taking the value one if the user tweeted the keywords *Chinese/Asian American* and *we* in column 1, a dummy taking the value one if the user tweeted the keywords *Chinese/Asian American* but did not tweet any keywords that signal reported discrimination in column 2, a dummy taking the value one if the user tweeted assimilation content but did not tweet the keyword *feel* in column 3, and a dummy taking the value one if the user tweeted assimilation content based on the subset of assimilation sentences including the keyword *American* in column 4. Results are local polynomial estimates using March 17 as the cutoff, controlling for dummies for day of the week and months of the year. Standard errors are clustered by date in parentheses, and statistical significance is computed based on the robust P value. Different bandwidths on each side of the cutoff are derived under the MSE procedure using a linear polynomial and a uniform kernel.

Second, to mitigate the concern that our results are driven by specific sets of users, in Tables 4 and 5 we show the robustness of our results to a different subsample of the Chinese users. In Table 4, we present estimates excluding political and activist users. Despite a sensible sample reduction (around 40%), all our results still hold, with only small decreases in magnitudes. Our results are also robust (and the coefficient larger in magnitude) if we focus only on the sample of users that report political or activist keywords (see Table 5). These exercises suggest that the rise in assimilation behavior on March 17 is not driven by our initial set of users.

Table 4: Robustness Assimilation: RDD Estimates, Cutoff March 17

Sample	Excluding Users w/Political+Activist Keywords in Bio		
Dep. Var.	Chin./Asian Amer.	Assimilation	Blame CCP Topic
	(1)	(2)	(3)
RD_Estimate	0.0103	0.0218	0.0029
	(0.0032)	(0.0033)	(0.0007)
Robust P-value	0.0925	0.0000	0.0010
Observations Left	3542	3542	7
Observations Right	14168	10120	30
Polynomial Order	1	1	1
Band. Method	msetwo	msetwo	msetwo
Band. Left	7.000	7.000	7.000
Band. Right	27.901	19.349	29.300
Day and Month Dummies	✓	✓	✓

Notes: We consider the sample of tweets of Chinese users after March 9. In column 1, the unit of observation is the user-day and the dependent variable is a dummy taking the value 1 if the user tweeted the keywords *Chinese American* or *Asian American*. In column 2, the unit of observation is the user-day and the dependent variable is a dummy taking the value 1 if the user posted a tweet containing assimilation content. In column 3, the unit of observation is the day and the dependent variable is the average of the share of text on the topic "Blame CCP" computed within the sample of users who tweeted that day. Estimates rely on the subsample of users whose Twitter bio does not include keywords related to politics and activism. Results are local polynomial estimates using March 17 as the cutoff, controlling for dummies for days of the week and months of the year. Standard errors, clustered by date in columns 1–2 and robust in column 3, are reported in parentheses. Statistical significance is computed based on the robust P value. Different bandwidths on each side of the cutoff are derived under the MSE procedure using a linear polynomial and a uniform kernel.

Table 5: Robustness Assimilation: RDD Estimates, Cutoff March 17

Sample	Only Users w/Political+Activist Keywords in Bio		
Dep. Var.	Chin./Asian Amer.	Assimilation	Blame CCP Topic
	(1)	(2)	(3)
RD_Estimate	0.0281	0.0215	0.0073
	(0.0052)	(0.0073)	(0.0011)
Robust P-value	0.0126	0.0454	0.0000
Observations Left	2282	2282	7
Observations Right	4890	6194	35
Polynomial Order	1	1	1
Band. Method	msetwo	msetwo	msetwo
Band. Left	7.000	7.000	7.000
Band. Right	14.643	18.444	34.466
Day and Month Dummies	✓	✓	✓

Notes: We consider the sample of tweets of Chinese users after March 9. In column 1, the unit of observation is the user-day and the dependent variable is a dummy taking the value 1 if the user tweeted the keywords *Chinese American* or *Asian American*. In column 2, the unit of observation is the user-day and the dependent variable is a dummy taking the value 1 if the user posted a tweet containing assimilation content. In column 3, the unit of observation is the day and the dependent variable is the average of the share of text on the topic "Blame CCP" computed within the sample of users who tweeted that day. Estimates rely on the subsample of users whose Twitter bio includes keywords related to politics and activism. Results are local polynomial estimates using March 17 as the cutoff, controlling for dummies for days of the week and months of the year. Standard errors, clustered by date in columns 1–2 and robust in column 3, are reported in parentheses. Statistical significance is computed based on the robust P value. Different bandwidths on each side of the cutoff are derived under the MSE procedure using a linear polynomial and a uniform kernel.

Altogether, the findings in this section point to an increase in assimilation efforts of the Chinese minority as a response to a worse discriminatory environment, triggered by a political leader: in particular, Chinese minorities tended to assert their belonging to the majority group and to assert their distance from their Chinese origins.

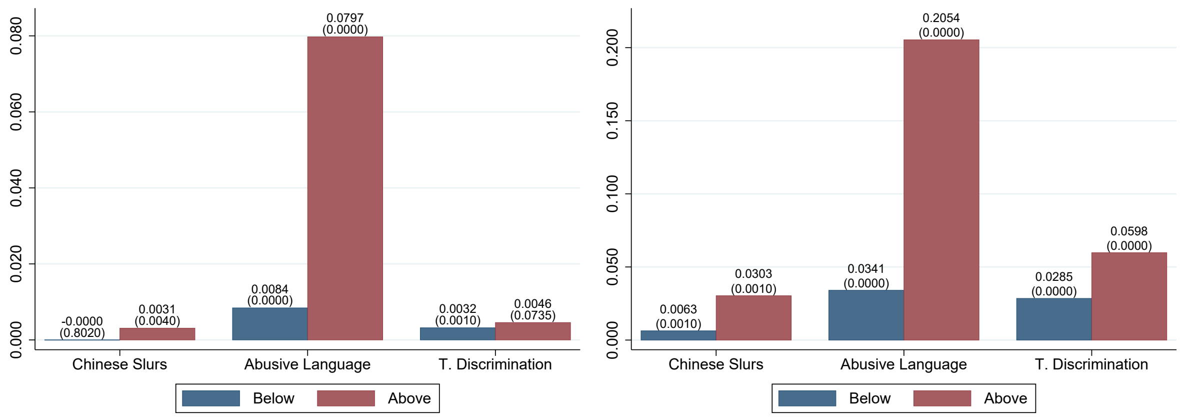
6 Heterogeneity of Results

We now consider possible heterogeneous effects stemming from preshock individual levels of discrimination and assimilation into U.S. society.

6.1 Heterogeneity of Discrimination Behavior of the "White Group"

We start by investigating whether the discrimination effects found within the White group depend on a generic initial individual propensity to discriminate, which we measure by calculating the users' share of tweets classified as generalized abusive language—considering *all* instances of abusive language, not only the ones directed against the Chinese community—from January 6 (the first day available in our sample) to February 17 (up to one month before Trump's discriminatory tweets). Panels (A) and (B) of Figure 8 report conditional estimates for all our discrimination proxies, comparing results from separate regressions on the subsample of White users with below versus above-median preshock share of generalized abusive language, using March 9 and 17, respectively, as the cutoff dates. Across all discrimination proxies, we find an increase in discrimination behavior for both groups as a result of both shocks; the increase triggered by both shocks is, however, larger and statistically significant for the set of users with a share of preshock generalized abusive language above the median.

Figure 8: Discrimination: Heterogeneity by Share of (general) Abusive Language Before February 17, 2020



(A) Cutoff: March 9, 2020 (B) Cutoff: March 17, 2020

Notes: Bars represent the point estimates of separately replicating even specifications in Panels A and B of Table 1 on the subset of users with a share of tweets with generalized abusive language below the median (blue) and above the median (maroon). The dependent variables are indicated on the x-axis. The magnitude of the coefficient and robust P value are reported above the bars, in parentheses.

6.2 Heterogeneity of Assimilation Behavior of the Chinese Minority

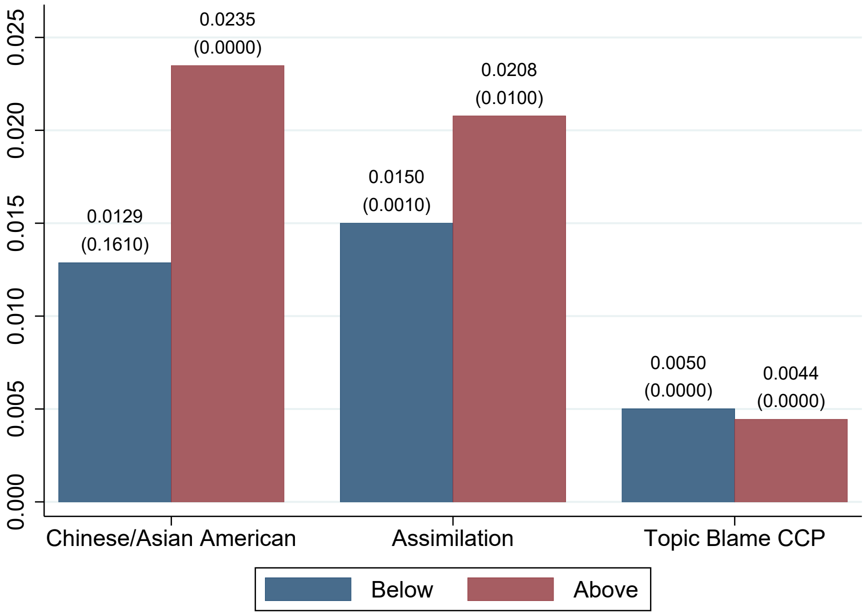
We now explore potential heterogeneous effects stemming from preshock individual levels of assimilation into U.S. society. In particular, we investigate whether the change in assimilation behavior after the Trump tweets found in the baseline analysis depends on Chinese users' share of Twitter friends located in the United States.[26] Running separate regressions for the subsamples of Chinese users with share of U.S. friends below versus above the median, Figure 9 reports conditional estimates for our assimilation proxies. We use March 17 as the cutoff date.[27] We find a rise in the assimilation behavior of both groups across all proxies; the increase in the likelihood that a user tweets the keywords "Chinese/Asian American" or assimilation content is much more pronounced for the set of users with an above-median share of U.S. friends.

Turning to the disidentification measure (average share of texts blaming the CCP), we find very similar coefficients between the two groups. Given the relevance of stigma for microlevel interactions (see Goffman, 2009, for a thorough discussion), this result plausibly reflects a first-order reaction to the association "virus-own ethnicity," which goes beyond user-specific initial levels of assimilation.

[26]This approach is in line with the empirical approach of Facchini et al. (2015), who find that immigrants with friends from the hosting country are more similar to natives.

[27]With respect to our preferred specification, all regressions additionally control for the share of friends for whom we could not retrieve the geographic location, but our results are unchanged if we remove this control variable.

Figure 9: Assimilation: Heterogeneity by Share of U.S. Friends Below and Above the Median



Notes: Bars represent the point estimate of separately replicating even specifications in Panel B of Table 2 additionally controlling for the share of friends who could not be geocoded. The results are based on the subset of users with share of U.S. friends below the median (blue) and above the median (maroon). The dependent variables are indicated on the x-axis. The magnitude of the coefficient and robust P value are reported above the bars, in parentheses.

7 Conclusion

Immigrants' assimilation into the hosting country has always been a central issue in the United States and other major receiving countries. Minority group members don't make assimilation decisions in isolation—these decisions often result from dynamic interactions with the majority group and with political leaders in a context with varying levels of discrimination.

This paper provides a unified framework to study the interactions between members of a discriminated minority, members of the majority group, and political leaders, from both a theoretical and an empirical point of view.

Theoretically, we set up a dynamic discrete-choice model in which forward-looking agents make discrimination and assimilation choices, and we analytically characterize the effect of a (temporary) exogenous shock on the return to discriminatory actions.

Empirically, we exploit novel Twitter data to study the discrimination and assimilation behavior of White Americans and the Chinese American community in the United States. To do this, we leverage two shocks to the discriminatory environment: the COVID-19 outbreak on March 9, 2020, and, one week later, then-President Trump issuing a tweet containing the phrase "Chinese Virus."

Three major results stand out. First, we show that White American Twitter users tended to discriminate more after both shocks, but the effect was stronger after Trump's discriminatory tweet. Second, we find that Chinese Twitter users significantly responded to the rise in discrimination following Trump's tweets by more forcefully asserting their American identity and by increasingly distancing themselves from the Chinese Communist Party. Third, these two sets of results are generally stronger when we narrow our focus to users with higher preshock levels of discrimination and assimilation.

These findings on the U.S. context suggest that minorities may react by assimilating more after the receiving country's discriminatory environment worsens. These results are in line with the qualitative findings of Kibria (2000), which emphasize the perceived need of Chinese (and indeed all Asian) Americans to counteract the strong connotation of "foreignness" associated with the Asian race. Future empirical work should study how other discriminatory shocks affect the behavior of other minorities.

References

Abdelgadir, A. and V. Fouka (2020). Political secularism and muslim integration in the west: Assessing the effects of the french headscarf ban. *American Political Science Review 114*(3), 707–723.

Abramitzky, R., L. Boustan, and K. Eriksson (2020). Do immigrants assimilate more slowly today than in the past? *American Economic Review: Insights 2*(1), 125–41.

Acemoglu, D., T. A. Hassan, and A. Tahoun (2017, 08). The Power of the Street: Evidence from Egypt's Arab Spring. *The Review of Financial Studies 31*(1), 1–42.

Acemoglu, D. and M. K. Jensen (2015). Robust Comparative Statics in Large Dynamic Economies. *Journal of Political Economy 123*(3), 587–640.

Adena, M., R. Enikolopov, M. Petrova, V. Santarosa, and E. Zhuravskaya (2015). Radio and the rise of the nazis in prewar germany. *The Quarterly Journal of Economics 130*(4), 1885–1939.

Aguirregabiria, V. and P. Mira (2007). Sequential Estimation of Dynamic Discrete Games. *Econometrica 75*(1), 1–53.

Aguirregabiria, V. and P. Mira (2010, May). Dynamic discrete choice structural models: A survey. *Journal of Econometrics 156*(1), 38–67.

Akerlof, G. A. and R. E. Kranton (2000, 08). Economics and Identity*. *The Quarterly Journal of Economics 115*(3), 715–753.

Anderson, D. M., B. Crost, and D. I. Rees (2020). Do economic downturns fuel racial animus? *Journal of Economic Behavior & Organization 175*, 9–18.

Anderson, R. W., N. D. Johnson, and M. Koyama (2017). Jewish persecutions and weather shocks: 1100–1800. *The Economic Journal 127*(602), 924–958.

Ash, E., D. Chen, and S. Naidu (2021). Ideas have consequences: The impact of Law and Economics on American Justice.

Aspachs-Bracons, O., I. Clots-Figueras, J. Costa-Font, and P. Masella (2008). Compulsory language educational policies and identity formation. *Journal of the European Economic Association 6*(2-3), 434–444.

Avitabile, C., I. Clots-Figueras, and P. Masella (2013). The effect of birthright citizenship on parental integration outcomes. *The Journal of Law and Economics 56*(3), 777–810.

Bailey, M., D. M. Johnston, M. Koenen, T. Kuchler, D. Russell, and J. Stroebel (2022). The Social Integration of International Migrants: Evidence from the Networks of Syrians in Germany. Working Paper w29925, National Bureau of Economic Research.

Barberá, P., A. Casas, J. Nagler, P. J. Egan, R. Bonneau, J. T. Jost, and J. A. Tucker (2019). Who leads? who follows? measuring issue attention and agenda setting by legislators and the mass public using social media data. *American Political Science Review 113*(4), 883–901.

Becker, S. O. and L. Pascali (2019). Religion, division of labor, and conflict: Anti-semitism in germany over 600 years. *American Economic Review 109*(5), 1764–1804.

Bénabou, R. and J. Tirole (2011, May). Identity, Morals, and Taboos: Beliefs as Assets *. *The Quarterly Journal of Economics 126*(2), 805–855.

Bergin, J. and D. Bernhardt (1992). Anonymous sequential games with aggregate uncertainty. *Journal of Mathematical Economics 21*(6), 543–562.

Bergin, J. and D. Bernhardt (1995, October). Anonymous sequential games: Existence and characterization of equilibria. *Economic Theory 5*(3), 461–489.

Bertrand, M. and E. Duflo (2017). Field experiments on discrimination. *Handbook of economic field experiments 1*, 309–393.

Bisin, A., E. Patacchini, T. Verdier, and Y. Zenou (2008). Are muslim immigrants different in terms of cultural integration? *Journal of the European Economic Association 6*(2-3), 445–456.

Bisin, A., E. Patacchini, T. Verdier, and Y. Zenou (2011). Formation and persistence of oppositional identities. *European Economic Review 55*(8), 1046–1071.

Bisin, A. and G. Tura (2019). Marriage, fertility, and cultural integration in italy. Technical report, National Bureau of Economic Research.

Bleakley, H. and A. Chin (2010). Age at arrival, english proficiency, and social assimilation among us immigrants. *American Economic Journal: Applied Economics 2*(1), 165–92.

Blei, D. M., A. Y. Ng, and M. I. Jordan (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research 3*(Jan), 993–1022.

Bursztyn, L., G. Egorov, R. Enikolopov, and M. Petrova (2019). Social media and xenophobia: evidence from russia. Technical report, National Bureau of Economic Research.

Bursztyn, L., G. Egorov, and S. Fiorin (2020). From extreme to mainstream: The erosion of social norms. *American economic review 110*(11), 3522–48.

Cagé, J., A. Dagorret, P. A. Grosjean, and S. Jha (2020). Heroes and villains: The effects of combat heroism on autocratic values and nazi collaboration in france.

Cattaneo, M. D., N. Idrobo, and R. Titiunik (2018). *A practical introduction to regression discontinuity designs: Volume II.* mimeo.

Clots-Figueras, I. and P. Masella (2013). Education, language and identity. *The Economic Journal 123*(570), F332–F357.

Davidson, T., D. Warmsley, M. Macy, and I. Weber (2017). Automated Hate Speech Detection and the Problem of Offensive Language. In *Proceedings of the Eleventh International AAAI Conference on Web and Social Media, ICWSM 2017*, pp. 4.

DellaVigna, S., R. Enikolopov, V. Mironova, M. Petrova, and E. Zhuravskaya (2014). Cross-border media and nationalism: Evidence from Serbian radio in Croatia. *American Economic Journal: Applied Economics 6*(3), 103–32.

Devakumar, D., G. Shannon, S. S. Bhopal, and I. Abubakar (2020). Racism and discrimination in covid-19 responses. *The Lancet 395*(10231), 1194.

Dipoppa, G., G. Grossman, and S. Zonszein (2021). Locked down, lashing out: Situational triggers and hateful behavior towards minority ethnic immigrants. *Available at SSRN: https://ssrn.com/abstract=3789339*.

Dippel, C. and S. Heblich (2021). Leadership in social movements: Evidence from the" forty-eighters" in the civil war. *American Economic Review 111*(2), 472–505.

Dustmann, C. and I. Preston (2001). Attitudes to ethnic minorities, ethnic context and location decisions. *The Economic Journal 111*(470), 353–373.

Echenique, F. (2005). A short and constructive proof of tarski's fixed-point theorem. *International Journal of Game Theory 33*, 215–218.

Enikolopov, R., M. Petrova, and K. Sonin (2018, January). Social media and corruption. *American Economic Journal: Applied Economics 10*(1), 150–74.

Facchini, G., E. Patacchini, and M. F. Steinhardt (2015). Migration, friendship ties, and cultural assimilation. *The Scandinavian Journal of Economics 117*(2), 619–649.

Fernández, R. and A. Fogli (2009). Culture: An empirical investigation of beliefs, work, and fertility. *American economic journal: Macroeconomics 1*(1), 146–77.

Fouka, V. (2019). How do immigrants respond to discrimination? The case of Germans in the US during World War I. *American Political Science Review 113*(2), 405–422.

Fouka, V. (2020a). Backlash: The unintended effects of language prohibition in us schools after world war i. *The Review of Economic Studies 87*(1), 204–239.

Fouka, V. (2020b, January). Backlash: The Unintended Effects of Language Prohibition in U.S. Schools after World War I. *The Review of Economic Studies 87*(1), 204–239.

Fouka, V., S. Mazumder, and M. Tabellini (2022). From Immigrants to Americans: Race and Assimilation during the Great Migration. *The Review of Economic Studies 89*(2), 811–842.

Founta, A.-M., C. Djouvas, D. Chatzakou, I. Leontiadis, J. Blackburn, G. Stringhini, A. Vakali, M. Sirivianos, and N. Kourtellis (2018). Large scale crowdsourcing and characterization of twitter abusive behavior. In *11th International Conference on Web and Social Media, ICWSM 2018*. AAAI Press.

Giuliano, P. (2007). Living arrangements in western europe: Does cultural origin matter? *Journal of the European Economic Association 5*(5), 927–952.

Goffman, E. (2009, November). *Stigma: Notes on the Management of Spoiled Identity*. Simon and Schuster.

Gould, E. D. and E. F. Klor (2016a). The long-run effect of 9/11: Terrorism, backlash, and the assimilation of Muslim immigrants in the west. *The Economic Journal 126*(597), 2064–2114.

Gould, E. D. and E. F. Klor (2016b). The Long-run Effect of 9/11: Terrorism, Backlash, and the Assimilation of Muslim Immigrants in the West. *The Economic Journal 126*(597), 2064–2114.

Griffiths, T. L. and M. Steyvers (2004, April). Finding scientific topics. *Proceedings of the National Academy of Sciences 101*(suppl 1), 5228–5235.

Grosfeld, I., S. O. Sakalli, and E. Zhuravskaya (2020). Middleman minorities and ethnic violence: anti-jewish pogroms in the russian empire. *The Review of Economic Studies 87*(1), 289–342.

Grosjean, P. A., F. Masera, and H. Yousaf (2020). Whistle the racist dogs: Political campaigns and police stops. *UNSW Business School Research Paper Forthcoming*.

Guirkinger, C., J.-P. Platteau, and Z. Wahhaj (2021). Behind the veil of cultural persistence: Marriage and divorce in a migrant community.

Halberstam, Y. and B. Knight (2016). Homophily, group size, and the diffusion of political information in social networks: Evidence from twitter. *Journal of Public Economics 143*, 73–88.

Hansen, S., M. McMahon, and A. Prat (2018, May). Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach*. *The Quarterly Journal of Economics 133*(2), 801–870.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning* (Second edition ed.). Springer.

Holshue, M. L., C. DeBolt, S. Lindquist, K. H. Lofy, J. Wiesman, H. Bruce, C. Spitters, K. Ericson, S. Wilkerson, A. Tural, et al. (2020). First case of 2019 novel coronavirus in the united states. *New England Journal of Medicine*.

Hong, L. and B. D. Davison (2010, July). Empirical study of topic modeling in Twitter. In *Proceedings of the First Workshop on Social Media Analytics*, SOMA '10, New York, NY, USA, pp. 80–88. Association for Computing Machinery.

Hopenhayn, H. A. and E. C. Prescott (1992). Stochastic Monotonicity and Stationary Distributions for Dynamic Economies. *Econometrica 60*(6), 1387–1406.

Jaschke, P., S. Sardoschau, and M. Tabellini (2022, August). Scared Straight? Threat and Assimilation of Refugees in Germany. Working Paper 30381, National Bureau of Economic Research, 1050 Massachusetts Avenue, Cambridge, MA 02138.

Jedwab, R., N. D. Johnson, and M. Koyama (2019). Negative shocks and mass persecutions: evidence from the black death. *Journal of Economic Growth 24*(4), 345–395.

Jovanovic, B. and R. W. Rosenthal (1988). Anonymous sequential games. *Journal of Mathematical Economics 17*(1), 77–87.

Kertz, R. P. and U. Rösler (2000, January). Complete lattices of probability measures with applications to martingale theory. *Game theory, optimal stopping, probability and statistics 35*, 153–178.

Kibria, N. (2000, March). Race, Ethnic Options, and Ethnic Binds: Identity Negotiations of Second-Generation Chinese and Korean Americans. *Sociological Perspectives 43*(1), 77–95.

Lee, D. S. and D. Card (2008). Regression discontinuity inference with specification error. *Journal of Econometrics 142*(2), 655–674.

Light, B. and G. Y. Weintraub (2022). Mean field equilibrium: Uniqueness, existence, and comparative statics. *Operations Research 70*(1), 585–605.

Lu, R. and Y. Sheng (2020). From fear to hate: How the covid-19 pandemic sparks racial animus in the united states. *arXiv preprint arXiv:2007.01448*.

Manning, A. and S. Roy (2010). Culture clash or culture club? national identity in britain.

Mayda, A. M. (2006). Who is against immigration? a cross-country investigation of individual attitudes toward immigrants. *The review of Economics and Statistics 88*(3), 510–530.

Miao, J. (2006, May). Competitive equilibria of economies with a continuum of consumers and aggregate shocks. *Journal of Economic Theory 128*(1), 274–298.

Milgrom, P. and J. Roberts (1994). Comparing Equilibria. *The American Economic Review 84*(3), 441–459.

Milgrom, P. and C. Shannon (1994). Monotone Comparative Statics. *Econometrica 62*(1), 157–180.

Müller, K. and C. Schwarz (2020). From hashtag to hate crime: Twitter and anti-minority sentiment. *Available at SSRN 3149103*.

Müller, K. and C. Schwarz (2021). Fanning the flames of hate: Social media and hate crime. *Journal of the European Economic Association 19*(4), 2131–2167.

NPR (2020). Coronavirus: All 50 states report cases; south america has nearly 1,000 cases. Accessed Sep. 9, 2021 [Online], available at https://www.npr.org/sections/health-shots/2020/03/17/817096232/coronavirus-radical-change-to-life-as-covid-19-reaches-152-countries?t=1631182083547.

Prummer, A. and J.-P. Siedlarek (2017). Community leaders and the preservation of cultural traits. *Journal of Economic Theory 168*, 143–176.

Rust, J. (1988, September). Maximum likelihood estimation of discrete control processes. *SIAM Journal on Control and Optimization 26*(5), 1006–1024.

Schmidt, A. L., F. Zollo, M. Del Vicario, A. Bessi, A. Scala, G. Caldarelli, H. E. Stanley, and W. Quattrociocchi (2017). Anatomy of news consumption on facebook. *Proceedings of the National Academy of Sciences 114*(12), 3035–3039.

Serfozo, R. (1982). Convergence of lebesgue integrals with varying measures. *Sankhyā: The Indian Journal of Statistics, Series A*, 380–402.

Shayo, M. (2020). Social identity and economic policy. *Annual Review of Economics 12*(1), 355–389.

Sunstein, C. R. (2018). *# Republic: Divided democracy in the age of social media.* Princeton University Press.

Tahmasbi, F., L. Schild, C. Ling, J. Blackburn, G. Stringhini, Y. Zhang, and S. Zannettou (2021). "go eat a bat, chang!": On the emergence of sinophobic behavior on web communities in the face of covid-19. In *Proceedings of the Web Conference 2021*, pp. 1122–1133.

Topkis, D. M. (1998). *Supermodularity and Complementarity.* Princeton University Press.

Van Zandt, T. and X. Vives (2007, May). Monotone equilibria in Bayesian games of strategic complementarities. *Journal of Economic Theory 134*(1), 339–360.

Verdier, T. and Y. Zenou (2018, May). Cultural leader and the dynamics of assimilation. *Journal of Economic Theory 175*, 374–414.

Vidgen, B., S. Hale, E. Guest, H. Margetts, D. Broniatowski, Z. Waseem, A. Botelho, M. Hall, and R. Tromble (2020, November). Detecting East Asian Prejudice on Social Media. In *Proceed-*

ings of the Fourth Workshop on Online Abuse and Harms, Online, pp. 162–172. Association for Computational Linguistics.

Voigtländer, N. and H.-J. Voth (2012). Persecution perpetuated: the medieval origins of anti-semitic violence in nazi germany. *The Quarterly Journal of Economics 127*(3), 1339–1392.

Waseem, Z. and D. Hovy (2016, June). Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter. In *Proceedings of the NAACL Student Research Workshop*, San Diego, California, pp. 88–93. Association for Computational Linguistics.

White, A. I. (2020). Historical linkages: epidemic threat, economic risk, and xenophobia. *The Lancet 395*(10232), 1250–1251.

Yan, X., J. Guo, Y. Lan, and X. Cheng (2013, May). A biterm topic model for short texts. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13, New York, NY, USA, pp. 1445–1456. Association for Computing Machinery.

Yanagizawa-Drott, D. (2014). Propaganda and conflict: Evidence from the Rwandan genocide. *The Quarterly Journal of Economics 129*(4), 1947–1994.

Ziems, C., B. He, S. Soni, and S. Kumar (2020, May). Racism is a Virus: Anti-Asian Hate and Counterhate in Social Media during the COVID-19 Crisis. *arXiv:2005.12423 [physics]*.