Ideas Mobilize Leaders: The Diffusion of Communist Ideology in China*

Ying Bai† Ruixue Jia‡ Runnan Wang §

January 27, 2024

Abstract

This paper examines the importance of ideas in motivating leaders, specifically analyzing how exposure to Communist ideology influenced individuals's decision to join the Chinese Communist Party (CCP) during its formative stage. We focus on cadets at the Whampoa Military Academy, who played crucial roles in China's major 20th-century conflicts. Our identification strategy leverages the locality-time-content variation in the circulation of the *New Youth* magazine—the primary platform promoting Communism after the Treaty of Versailles in 1919—and the variation in an individual's location over time. Comparing cadets living in a locality with access to post-1919 *New Youth* against those who lived in the same locality but missed this channel, we find that the former were significantly more likely to join the CCP. Additionally, ideology exposure appeared consequential in subsequent conflicts. In contrast, family background did not predict party choice of these political pioneers.

*We thank Sam Bazzi, Ernesto Dal Bo, Harris Doshay, Fred Finan, Matthew Gentzkow, Saad Gulzar, Yasheng Huang, Debin Ma, Suresh Naidu, Nancy Qian, Marvin Suesse, Shaoda Wang, Andrew Walder, Chenggang Xu, Guo Xu, Yang Yao, David Yang, Noam Yuchtman, Enying Zheng, Fabrizio Zilibotti and seminar/conference participants at Academia Sinica, Berkeley, Call to Service Conference, Columbia, CUHK-Shenzhen, Harvard Business School, HKU 1st Annual Conference on China and the Global Economy, National School of Development (PKU), NBER China meeting, Northwestern-U Chicago Development and Political Economy Conference, New Economic History of China Conference, PKU-Tsinghua conference on Chinese economy, Quantitative History Webinar, Stanford, Trinity College Dublin, and UCSD for their comments.

†Department of Economics, Chinese University of Hong Kong; ybai@cuhk.edu.hk

‡UCSD, CEPR and NBER; rxjia@ucsd.edu

§Department of Economics, Chinese University of Hong Kong; rwang@link.cuhk.edu.hk

1 Introduction

Leaders are crucial in directing collective actions such as wars and revolutions, which have the potential to change the course of history. This study investigates the factors that motivate leaders, particularly in the initial, uncertain, and perilous stages of social and political movements. While the current economics literature provides insightful analyses of collective action motivations, it primarily highlights the importance of interests and often neglects the motivations of leaders.[1] In this paper, we shift the focus from interests to the power of ideas—a hypothesis that has yet to be extensively empirically validate—by examining the influence of Communist ideology's spread in China. We aim to illustrate how the dissemination of ideas can influence the decisions of political pioneers, potentially shaping the fate of a nation.

The spread of Communist ideology was one of the most significant political events of the 20 century: For a large part of the century, approximately one-third of humanity lived in Communist countries. Yet, little is known about how such an ideology got diffused across such vastly different contexts, ranging from Germany and Italy in the west to China and Indonesia in the east. China, during the 1910s–1920s, provides a useful context for understanding Communism's spread via the printing press. As well recognized by abundant historical narratives (discussed in Section 2), the most prominent press in spreading Communist ideology is the *New Youth* magazine (1915–1922), a leading literary magazine founded in 1915 that spearheaded the New Culture Movement (1915–1923) but turned to promoting Communism after the Treaty of Versailles in 1919.[2]

Identifying the impact of ideology diffusion, however, is challenging. First, it is difficult to differentiate ideological exposure from other confounding factors. For instance, the circulation of the *New Youth* magazine across localities would, in most cases, be associated with locality size or urbanization rates. Second, it is similarly hard to determine the direct population that was influenced by the ideology. In a setting like the one being discussed in this paper, where over 80% of the population were illiterate and printing press was the primary form of media, it is conceivable that only a small share of the population could be influenced. Although these individuals might go

[1] Influential research categorizes these interests into concepts of greed and grievances (Fearon and Laitin, 2003; Collier and Hoeffler, 2004). Extensive empirical studies have employed varied measures to represent these incentives, as summarized in Blattman and Miguel (2010). Additionally, social factors such as signaling one's image are also part of interests, as reviewed in Cantoni et al. (2023).

[2] Even though China was on the winning side of WWI, the Treaty of Versailles allowed Japan to retain territories in Shandong that had been surrendered to Germany. Such actions of Western powers undermined the moral prestige of the West in China and directed the leadership of the *New Youth* towards Russia for enlightenment. See more on the background in Section 2.

on to become pioneers in the coming Communist revolution, it is challenging to identify them with certainty.

To address these challenges, we leverage a specific and significant context: the choice of individuals at the Whampoa Military Academy to become a Chinese Communist Party (CCP) member during the CCP's formative stage in the 1920s. The Whampoa Military Academy, established in 1924, produced commanders and legendary graduates who would subsequently go on to fight China's most important wars of the 20 century, including two civil wars between the CCP and the Kuomintang (KMT) as well as the anti-Japanese war. It remains to be understood what motivated these individuals to become revolution leaders.

This setting presents four key advantages for assessing the impact of ideological exposure. First, the initial cohort at the Whampoa Military Academy consisted of 689 individuals who provided comprehensive data regarding their family background, life histories, and even the references they utilized when applying to Whampoa. Given the subsequent importance of these individuals, we can also gather information on their future political trajectories. This rich biographical information enables a thorough examination of individual-level political choices within a significant historical context, a task that has posed challenges for researchers in the past. Additionally, we incorporate data from an extra 700 cadets in subsequent cohorts to verify the consistency of our findings across different groups. Second, studying a group of individuals who shared the same environment at Whampoa allows for more accurate comparisons. Third, the introduction of Communist ideas to China in the early 1920s renders the influence of ideas more discernible compared to scenarios where certain ideas have a longer historical presence. Lastly, the availability of detailed data on individual mobility, the spread of the *New Youth* magazine, and archival information on other relevant periodicals and organizations provides valuable resources for identification and interpretation purposes.

We use two variations in our identification strategy: time-varying geographical variation in the circulation of the *New Youth* and its content change before and after 1919, and time-varying location of individuals due to education or employment. We compare cadets who were living in a locality with access to the post-1919 *New Youth* against those who were in the same locality but missed it due to the timing of their life events. By controlling for locality fixed effects, we can address the challenge that ideological exposure is correlated with other locality characteristics. Our analyses use two definitions of localities—prefectures and counties–for comparison and robustness.

We show that, conditional on locality fixed effects, individuals with or without access to post-1919 *New Youth* are comparable in terms of a rich set of personal characteristics, including family socioeconomic status, education, the political identities of their references, age, numbers of siblings and children, marriage, and even behavioral traits such as smoking and drinking. This

pattern suggests that an individual's location is primarily driven by economic concerns (especially education and jobs) rather than the presence of the post-1919 *New Youth*. Furthermore, we confirm that selection bias into our Whampoa sample isn't a major concern: once locality-related fixed effects are accounted for, there's no discernible correlation between exposure to post-1919 *New Youth* and either the inclusion in our sample of Whampoa cadets or their specific characteristics.

Our main finding is that exposure to the *New Youth* at the prefecture during 1919–1922 increases an individual's probability of joining the CCP by 11.7 percentage points compared to individuals who resided in the same locality but missed this exposure. The magnitude becomes 17.2 percentage points when we zoom into county-level exposure. These estimates imply a persuasion rate of at least 18.9%.[3] This relatively high persuasion rate is partly because our studied sample is a group of educated young radicals who were actively making a decisive choice regarding party affiliation. In contrast, the other personal characteristics, except for the political identity of one's references, cannot predict the decision to join the CCP and the KMT and do not offer strong heterogeneous impact for ideology exposure. Of course, these auxiliary findings do not imply that Whampoa cadets were similar to the general population; in fact, they were distinct (refer to Section 2.4 for characteristics of the cadets). Our focus is to understand the choices among these cadets, who would later become leaders of opposing sides in major civil wars.

Two primary concerns of our main finding are whether the *New Youth* entered certain localities due to potential demand and whether individuals relocated to those localities in response to the availability of post-1919 *New Youth*. We should note that our estimate is similar even if controlled for the interactions between prefecture fixed effects and the post-1919 dummy, which allows us to then control for possible changes in the demand for Communism in different prefectures before and after 1919. Moreover, we take a closer look at possible endogeneity concerns regarding the circulation locations of the *New Youth* and the relocation of individuals. We find that exposure to the *New Youth* during 1915–1918 had no similar influence on the CCP choice, suggesting that our finding is not driven by the endogenous entry of the *New Youth*. Additionally, our estimate is similar when restricting exposure to the localities that *always* had access to the *New Youth*, which helps eliminate the possible entry concerns of *New Youth* after 1919.

We conduct two sets of analysis to check the relevance of individuals' sorting toward post-1919 *New Youth*. First, to capture the location preferences for an individual, we construct placebo locations of an individual by assuming that he were in his previous or following location based on his life event records. We use these placebo locations to calculate placebo exposures to post-1919 *New Youth* and find that these placebo exposures cannot explain our finding either. Second, we separate

[3]We also explore heterogeneity across localities and obtain an upper bound of 29.2%. DellaVigna and Gentzkow (2010) summarize persuasion rates across different settings, ranging from 0.7% to 30%. Our estimates are sizable compared with those in the literature.

our comparison group who had resided in prefectures ever with post-1919 *New Youth* but missed it into several cohorts, including those who arrived before 1919, during 1919–22, and in 1923. Comparing with these various cohorts, we find similar estimates to our baseline results, implying that it is difficult to use individuals' sorting to explain our findings. We also address concerns regarding possible measurement error at both the locality and individual levels. Additionally, we extend our analysis to include potential spillover effects and another 700 selected cadets from the second to fourth cohorts, and our findings remain robust.

We focus on post-1919 *New Youth* exposure because it is widely recognized as the main channel for Communist ideology diffusion (see qualitative evidence in Section 2). It is certainly possible, however, for our studied individuals to read other magazines or be exposed to activist groups in this period, some of which might also help spread Communist ideology. Because of this, we conduct three sets of analyses to help interpret post-1919 *New Youth* exposure. First, based on archival information on pro-Communism periodicals that existed during 1919–1922, we are able to characterize a network by examining whether these magazines shared editors or editorial offices. This network reveals that the *New Youth* was the most central magazine and its editors often would be instrumental in helping co-found other magazines. Comparing the exposure to post-1919 *New Youth* to that of two other pro-Communism magazines, we find that few would have been exposed to the other magazines without being exposed to post-1919 *New Youth*. Estimates are similar regardless of whether we further consider these additional periodicals. Second, we use archival information on activist groups in this era and find our result still holds in localities without these groups. Third, we leverage the timing of joining the CCP and information on potential social interactions in locality of exposure to show that it would be difficult to use factors such as local CCP recruitment to account for our main finding. Based on these analyses, we interpret post-1919 *New Youth* exposure as a reasonable proxy for the diffusion of Communism ideology in a locality.

In the latter part of our analyses, we investigate whether individuals who were mobilized by Communist ideology exhibit different behavior in critical historical moments. This is a challenging yet important issue to study. To explore this, we examine the future political outcomes of our studied individuals during various political struggles in which many quit their parties and sacrificed their lives. We find suggestive evidence that those who were more influenced by Communist ideology exposure were less likely to quit the CCP and more likely to sacrifice their lives. This difference is primarily driven by the period from 1927 to 1936, during the first Civil War between the CCP and the KMT when the CCP was the weaker party and was under severe attacks by the KMT. These findings suggest that initial ideology exposure can work in conjunction with other factors of the revolution process to significantly influence an individual's political choices.

This paper demonstrates the importance of ideas in shaping the decisions of political leaders,

intersecting with multiple lines of research. An emerging body of work in political economy has documented the pivotal role of leadership in economic and political development (e.g., Jones and Olken, 2005; Dippel and Heblich, 2021; Cagé et al., 2023; Bai et al., 2023). However, our understanding of what drives these leaders, especially beyond material interests, remains limited. This is particularly pertinent in the context of political and social movements facing substantial uncertainty.

Our research also augments the increasingly rich literature on the impact of media in shaping political behaviors, which has documented the influence of various forms of media in a variety of contexts (see DellaVigna and Gentzkow (2010), Prat and Strömberg (2013), and Zhuravskaya et al. (2020) for reviews). For instance, influential research has revealed radio's role in fueling genocide (Yanagizawa-Drott, 2014) and its contribution to the rise of Nazism (Adena et al., 2015), as well as cinema's influence in perpetuating racial hatred (Ang, 2023). Most of the existing research focuses on macro-regional patterns. Our study contributes new individual-level insights, offering comparisons between idea exposure and socioeconomic background. We also find suggestive evidence that such an impact may partly determine the life and death of political pioneers at critical historical moments.

The role of ideas in shaping economic and political processes has been a subject of scholarly interest for a long time. Joel Mokyr, for instance, has highlighted the role of Enlightenment ideas in spurring the Industrial Revolution (Mokyr, 2005). Our study joins recent economic research examining the influence of ideological exposure on voting behavior (e.g. DellaVigna and Kaplan, 2007), public opinion (Cantoni et al., 2017), and policy-making (Ash et al., 2022; Spenkuch et al., 2023). We provide novel evidence on how Communist ideas influenced individual political choices, thereby enhancing our understanding of ideas' roles in significant historical moments.[4] Moreover, the impact of ideas is not confined to the Communist revolution. The influence of ideas in other major revolutions, such as the American and French Revolutions, has long been argued and debated (e.g., Bailyn, 1967).[5] Recent empirical studies are beginning to explore these influences in greater depth (e.g., Jha and Wilkinson, 2023; Ottinger and Rosenberger, 2023). Our findings suggest that ideas may be particularly relevant for the revolutionary leaders, particularly those who joined movements at their inception, a hypothesis that awaits exploration in other contexts.[6]

[4]Numerous historical studies on Communism exist, although they primarily focus on a few nations as case studies (Smith, 2014). Magness and Makovi (2023) document the Russian Revolution's central role in mainstreaming the ideas of Marx. In our setting, the Russian Revolution and the Treaty of Versailles facilitated the adoption of Communism in China.

[5]The Iranian Revolution may serve as another pertinent example (e.g., Shadmehr, 2017).

[6]In contrast, individuals who joined the CCP after it gained prominence likely differ from the early pioneers we study. Scholars have highlighted the role of nationalism against Japan in CCP's growth post-1937, illustrating the complex dynamics during different periods (Johnson, 1962, Chen and Kung, 2020).

2 Background and Data

2.1 The *New Youth* (1915–1922) and the Founding of the CCP (1921)

The period of 1910s–1920s, known as the "Chinese enlightenment" (Schwarcz, 1986), is exemplified by the New Culture Movement (1915–1923), the later stage of which is also known as the May Fourth Movement (1919–1923).[7] Arising out of a disillusionment with traditional Chinese culture after the failure of the Beiyang government (1912–1928) to address social and economic problems, many intellectuals joined the movement to promote a new culture based upon western ideals, particularly democracy and science. The movement was launched by the writers of the *New Youth*, a monthly magazine founded in 1915 by Chen Duxiu. Different from those who thought that democracy and constitutional government could be easily established in China, Chen argued that there must be a change in *thought* to support the constitutional government (Chen, 1916). He believed that there couldn't be a new political system without a new culture.

Following the emphasis on culture, the *New Youth* magazine initially focused on culture and literature, offering such debates as promoting written vernacular Chinese over classical Chinese. It became popular very quickly with its monthly circulation rising from 1,000 for the first issue to over 16,000 after 1917 (Chow, 1960). Over time, frustrated with the inadequacy of cultural change and influenced by the Russian Revolution, the magazine, led by Chen Duxiu and Li Dazhao, started considering more political action. The 1919 Treaty of Versailles marked a turning point, where Chinese intellectuals, disillusioned at the compromising of Chinese interests by Western powers, found the Russian way as a new role model.[8] In May 1919, the *New Youth* published a special issue on Communism. After that, the magazine became increasingly aligned with the Communist ideology and served as the fountainhead of Chinese Communism (Yoshihiro, 2012).

To track the ideological shift in the *New Youth* magazine, we analyze the distribution of article titles as depicted in Figure 1. Panel (a) highlights the mention of different nations, showing that France and Britain were initially prominent, reflecting a focus on Western ideals. However, a significant transformation occurred after 1919, with Russia becoming the most discussed nation, comprising 10–15% of annual articles. In Panel (b), the dominance of culture and literature topics initially gives way to political discussions on Communism and revolution from 1919–20 onwards. In Appendix A.1, we provide additional insights by examining the most frequently used words in titles. Notably, post-1919, words associated with Communism (e.g., labor, Russia, socialism,

[7] The May Fourth Movement is named after the May Fourth incident on May 4, 1919. On that day, students gathered in front of Tiananmen to protest the government's weak response to the Treaty of Versailles. These demonstrations sparked nationwide protests and contributed to the demand for political action.

[8] In sharp contrast with the Treaty of Versailles, the Karakhan Manifesto, issued in 1919 on behalf of the Soviet government, nullified unequal treaties forced upon China by Czarist Russia. Perhaps not surprisingly, Russia replaced the West as the symbol of justice and freedom in China.

Marx, and revolution) prominently emerge, indicating its relevance as a set of ideas centering around proletarian revolution and the concept of organized "unified action" (Dirlik, 1989).

The leaders of the *New Youth* were also among those who founded the CCP. In July 1921, the first National Congress of the CCP was held in Shanghai. It was here that Chen Duxiu was elected as the Chief Party Secretary, and following the event, the *New Youth* would then become a theoretical journal for Communism. The *New Youth* published its last issue in July 1922.[9] The *New Youth* published 54 issues in total during 1915–1922, which has served as an important source for studying the political intellectual history of modern China.

Personal Accounts. There is ample anecdotal evidence regarding the influential role of the *New Youth* magazine in shaping the ideology of the youth in the early 20th century, including memoirs from future leaders of the Chinese Communist Party (CCP). For instance, Mao Zedong was an avid reader of the magazine and made his first contribution to it in 1917. In his 1936 interview with Edgar Snow (Snow, 1937), Mao reminisced:

> "I began to read this magazine [*New Youth*] while I was a student in the normal school and admired the articles of Hu Shi and Chen Duxiu very much. They became for a while my models, replacing Liang Qichao and Kang Youwei."[10]

Deng Xiaoping provides another example of being influenced by the *New Youth* (more specifically, post-1919 *New Youth*). In 1926, while studying in Moscow, he wrote an autobiography reflecting on his journey towards becoming a socialist believer (Yang, 1994):

> "At that time [1922], my life was still considered romantic and not suitable for so-called "respectable" people. However, I felt that I was making progress because I began reading books and newspapers about socialism. The most influential works for me were the 8th and 9th volumes of the *New Youth* and a collection of socialist discussions. My work environment reinforced my belief in the words of Chen Duxiu and his colleagues, so whenever I heard people arguing with each other, I always sided with the socialist perspective."

Similar recognitions of the formative role of the *New Youth* can be found in life experiences of other Communist pioneers including Peng Dehuai in Hunan (Snow, 1937), Zhou En-lai in Tianjin (Snow, 1937), Zhang Guotao in Beijing (Zhang, 1971), Yun Daiying in Hubei (Yun, 1981), and Shi Cuntong in Zhejiang (Yeh, 2023). We summarize these personal accounts in Appendix A.2. These accounts reveal two patterns. First, the *New Youth* was influential across many regions in

[9] Although the original *New Youth* was stopped in 1922, the CCP continued to use the title for a new quarterly journal during 1923–1926. Because of the timing of our study, we do not focus on this latter period.

[10] Liang Qichao and Kang Youwei were influential intellectuals who proposed reforms (instead of revolution) of the imperial system.

China. Second, the *New Youth* served as a source of inspiration for other publications (which are explored in Section 4.2), and it was common for some young people to read the *New Youth* and its satellite publications.

Given the importance of the *New Youth*, few scholars studying the history of the CCP have missed mentioning it. As put by Meisner (1999), "it is difficult to overestimate the importance of the intellectuals who coalesced around the *New Youth*, for their writings molded the beliefs and attitudes of a whole generation of young students who were to achieve political prominence after the May Fourth incident of 1919 and who were to become the leaders of the modern Chinese revolution."

Other Magazines and Organizations. As reflected by the personal accounts in Appendix A.2, the post-1919 *New Youth* was the major but not the only channel for Communist ideology diffusion. Following in the footsteps of the *New Youth*, quite a few pro-Communism periodicals emerged, some of which were co-founded by the editors of *New Youth*. Moreover, there existed various activist groups that also helped spread Communism. In Section 4.2, we examine how these alternative channels contribute to our understanding of the role played by the *New Youth*.

2.2 The Whampoa Military Academy (1924–1927)

The Whampoa Military Academy was established in 1924 in Whampoa, Guangzhou, as proposed and supported by the Soviet Union. In 1924, under the demand of the Soviet Union, the KMT formed a political alliance with the CCP to fight against the incumbent Beiyang government and other regional warlords. In this era, known as the First United Front, the KMT was the dominant revolutionary party whereas the CCP was at its more nascent stage. While initially everyone at Whampoa was required to be a KMT member, one could choose to join the CCP if they hadn't before 1924.

However, the KMT-CCP collaboration ended in 1927, and a civil war broke out, lasting until 1937. After the breakup, one was only able to have one party identity. Afterward, during 1937–45, the KMT and the CCP formed a second political alliance, the Second United Front, to fight against Japan. After winning the anti-Japanese war, the KMT and the CCP fought in another civil war until the CCP won the mainland, and the KMT withdrew to Taiwan in 1949–1950.

Despite its relatively short presence, the Whampoa Military Academy played a profound role in Chinese modern history.[11] Whampoa concentrated the revolutionary military talents into training regiments that followed the Soviet Red Army model. The exigencies of the revolution limited the time available for training, lasting only half a year for most cohorts. The emphasis was on the

[11] After the KMT-CCP breakup, the KMT established the Republic of China Military Academy in Nanjing.

practical knowledge and skills required in the battlefield (Graff and Higham, 2012). Whampoa cadets would become commanders for both the KMT and the CCP and fight both the civil wars and the anti-Japanese war.[12]

Moreover, the Whampoa graduates have influenced both parties' policies and governance. For Chiang Kai-shek, the leader of the KMT, the "Whampoa clique" was a pivotal force for his military organization and governance (Taylor, 2009). For the CCP, when it built its first Red Army in 1927, most of its commanders were from Whampoa. In the following two decades, the CCP trained its army in the Whampoa way, which would finally serve in bringing the CCP victory.[13] In 1955, the CCP recognized the Ten Founding Marshals of the People's Republic of China, and five out of the ten had affiliations with Whampoa.

2.3 The Timeline

We present a timeline of significant events from the 1910s to 1949 in Figure 2. Our study specifically examines how exposure to Communist ideology through the *New Youth* magazine during 1919–1922 influenced individuals' decision to join the Chinese Communist Party (CCP) during its formative stage in the 1920s. As depicted in Figure 2(a), the number of CCP members initially remained small but experienced a remarkable growth in this period. At the 3rd National Congress of the CCP in June 1923, there were only 432 registered Communist members, but by the 5th National Congress in April 1927, the number had surged to over 57,000. We also explore the subsequent trajectories of the individuals in our study, spanning the 1920s to 1949, to gain insights into the extended impacts of ideology exposure.

Figure 2(b) presents the founding years of Communist parties in different countries, along with the year of the translation of the *Communist Manifesto*. This visualization reveals that the Russian Revolution spurred the emergence of many new Communist parties worldwide. Thus, the formation of the CCP is one among many examples influenced by both international politics and domestic tensions. It is noteworthy that the Chinese translation of the *Communist Manifesto* was published in 1920, relatively later than in many other countries. This highlights that Communist ideas were relatively new to China and aligns synchronously with our timeline of 1919–1920.

[12]For instance, legendary commanders on the KMT side include Chen Cheng, Du Yuming, Xue Yue, Hu Zongnan, Hu Lien, and Guan Linzheng, whereas the CCP side includes Lin Biao, Xu Xiangqian, Liu Zhidan, and Chen Geng.

[13]In a 1937 interview with James Bertram, Mao Zedong emphasized that the Red Army had inherited and advanced the traditions and practices of the Whampoa.

2.4 Data

We present the summary statistics in Appendix A.3 and the detailed data construction process in Appendix A.4–A.5. Below, we describe our key variables.

Whampoa Cadets. Our main analysis centers on the first cohort of the Whampoa Military Academy. To verify the applicability of our findings, we also incorporate data from selected cadets of the second through fourth cohorts for comparative assessment.

The initial cohort consisted of 706 individuals, with 17 missing information about their background.[14] Our analysis focuses on the remaining 689 individuals for whom we have information. Out of these, 530 were directly recruited by Whampoa and were required to provide comprehensive details about their family background and life history at the time of enrollment in 1924. At Whampoa, they were assigned to Teams 1–4. Appendix A.4 provides an example of the life history records based on the *Detailed Survey for the First to Fourth Teams of the First Cohort of the Military Academy* (Central Executive Committee of the Kuomintang, 1924). The remaining 159 cadets were recruited from a regional military school in Guangzhou. We obtain information on the background of all 689 individuals from the work of Chen Yuhuan, a historian and Whampoa expert (Chen, 2017a). In our analyses, we consider the entire sample of 689 individuals and present results specifically for Teams 1–4. Appendix A.5 provides an example of our coding system for tracking life events.

It is important to note that not all individuals recorded their life events on a year-by-year basis. The datasets contain two types of missing information. First, some individuals omitted the end year of their life events. In such cases, we assume that the life event lasted until the start of the next recorded event. Second, some individuals reported certain life events but did not specify the year. In these instances, we infer the years based on their birth year, the end year of available records, and knowledge about the school system between 1912 and 1922 (Fan and Popkewitz, 2020).[15] To check the validity of our inference, we present results by pretending that we do not know the specific years and applying the same inference method to all our sample. We also conduct checks regarding the precision of life events by examining how the number of records affects our findings.

While the first cohort at Whampoa was surveyed systematically, the other cohorts were not. However, historical research has yielded biographical information on a substantial number of cadets from the second to fourth cohorts, albeit in a non-systematic manner. To supplement our analysis of the first cohort, we digitize the biographical data of 700 cadets from the second to fourth cohorts, based on Chen (2017b), Chen (2017c), and Chen (2018).

[14] These individuals either did not appear at Whampoa or potentially left before graduation.

[15] This system is known as the *Renzi Guichou* school system, implemented during 1912 (the year of *Renzi* in the Chinese calendar) and 1913 (the year of *Guichou*). This system remained in effect until 1922.

CCP Membership and Future Political Outcomes. Our primary outcome variable of interest is CCP membership. We gather data on the political party affiliation and subsequent outcomes of the individuals in our sample. This includes information on whether they chose to quit their party or if they died in political struggles, particularly between the KMT and the CCP. We code these variables based on information provided by Chen (2017a), and we validate the data using Wikipedia and Baidu. While it is possible for individuals to have concealed their CCP membership in the 1920s, this is not a concern for our study as our source on CCP membership is based on information disclosed later.

During the 1920s, the CCP was considerably weaker compared to the KMT: Only a relatively small proportion of our sample (17.1%) joining the CCP. When examining the timeline of CCP membership (presented in Appendix A.6), we observe that all participants who joined the CCP did so during its early years (1921–1928), with 77% of them joining during or after the Whampoa period (i.e., 1924–1928). The small number of individuals joining the CCP before the Whampoa period suggests limited access to party organizations during that time, as evidenced by the mere 432 registered CCP members across the country by 1923.

Our study focuses on a critical moment in history where the cadets in our sample faced significant risks and uncertainties in their political careers. Specifically, 24.1% of them experienced death in future conflicts, and 16% chose to quit their respective parties, including instances of surrendering to the opposing party in the conflicts. These outcomes highlight the high stakes and volatility of the revolutionary era, suggesting that the political identity choices made by the cadets could have substantial consequences.

Personal Characteristics. We consider 13 variables of a Whampoa cadet's personal characteristics categorized into three groups, which together provides a clear portrait: These were a group of men in their early 20s in 1924 who had received middle-school or secondary-school education and came from families of the middle and lower strata.

(i) Education and socioeconomic background. As reported in Appendix A.3, 96% of the cadets had completed middle school and 61% attended secondary vocational schools. 11% had some college education or attended some college classes but usually did not hold college degrees. 54% came from families with some education or business tradition; 39%, 58%, and 3% ranked their family status as "poor", "middle-income", and "rich". We code this status as a categorical variable (1–3), which has a mean value of 1.63.

Alongside education and socioeconomic background, we know the political identity of the personal references that cadets used in their application to the Whampoa and can measure how many CCP references one had. This variable proxies whether one had direct or indirect connections with existing CCP members and can be considered as a crude measure of social networks. This variable

ranges from 0 to 5 and has a mean value of 0.18, which again implies that few could directly access CCP references by 1924.

(ii) Demographic information. The average age of the studied individuals was 23 in 1924. The average number of siblings they had is 3.4. 54% of them were married and 14% had children, reflecting marriage at an early age in this era.

(iii) Religion and behavior traits. 11% reported believing in some religion (mostly Confucianism and Buddhism). 5% reported a habit of smoking or drinking.

Exposures to the *New Youth* and Communist Ideology. Each year, the *New Youth* discloses information on its circulation locations in at least one issue, which we use to construct prefecture-year and county-year datasets on the availability of the *New Youth*. The variation in circulation is driven by both the expansion effort of the magazine and the censorship imposed by the Beiyang government. We present an example of this circulation data and a map of circulation locations over time in Appendix A.5 (Step 1). By 1920, the *New Youth* had reached 53 prefectures, covering 37.9% of China's population. See the maps in Appendix A.5 for the circulation locations. Notably, the *New Youth* was always available for a group of prefectures whereas there existed entry and exit in others. We leverage such a variation to examine whether the potential endogeneity concerns of the *New Youth* circulation matter for our findings.

We measure the *New Youth* exposure by analyzing the intersection between the locality-year circulation variation and the variation in an individual's life history. In Appendix A.5 (Step 2), we explain our life history coding process and present maps of cadets' locations over time. By 1924, the individuals we study had been to 160 prefectures, covering 80.0% of China's population.

Considering that A and B lived in the same locality, Person A could have been exposed because he happened to be there during the years the *New Youth* was available, whereas Person B could have not been present during the years it was available. We illustrate the intersection between cadets' locations and *New Youth* locations in our data in Appendix A.5 (Step 3). Further, our measure of Communist ideology exposure considers the content change of the *New Youth* before and after 1919. If an individual was living in a prefecture (or county) when post-1919 *New Youth* was available, we define him as being exposed to Communist ideology. By this design, we study an intention-to-treat effect. In addition to this dummy exposure indicator, we present the results using a continuous variable that calculates the number of exposure years.

3 Descriptive Evidence and Research Design

Descriptive Evidence. Did exposure to the *New Youth*, more specifically to post-1919 *New Youth*, have an impact on joining the CCP? We present two pieces of evidence from the raw data to motivate our empirical analyses.

First, we explore the temporal pattern by dividing individuals into an exposed group (those living in a prefecture with access to the *New Youth* after 1919) and a non-exposed group and comparing their probabilities of joining the CCP. We compare their probabilities of joining the CCP over time. As depicted in Figure 3(a), the difference in probabilities becomes positive after 1919–20, coinciding with the period when the *New Youth* actively promoted Communism. A similar pattern emerges when we define *New Youth* exposure at the county level, as illustrated in Figure 3(b).

Second, we examine the spatial pattern, which suggests the relevance of ideology exposure across different localities. Figure 4(a) shows the number of CCP members among the group with post-1919 *New Youth* exposure across prefectures, while Figure 4(b) shows the number among the group without post-1919 *New Youth* exposure. On average, the group with post-1919 *New Youth* exposure has a higher probability of joining the CCP in the raw data, with mean ratios of 19.3% and 11%, respectively.

Although these patterns offer suggestive evidence, their interpretation is challenging due to the correlation of post-1919 *New Youth* exposure with various other characteristics specific to each locality. Therefore, in our subsequent analysis, we employ a within-locality research design to address these concerns.

Estimation Strategy. Our estimation strategy takes into account the localities individuals had been to during the period of 1915–1922 by including locality fixed effects and focusing on the intersection between two key factors: (1) the variation in *New Youth* availability at the locality-year-content level and (2) the variation in an individual's life events at the locality-year level. We will also address potential endogeneity concerns related to each variation, such as by examining localities where the *New Youth* was consistently available. We begin with a baseline specification for prefecture-level exposure as follows:

$$CCP_i = \beta NYExposure_{i,p,1919-22} + \alpha NYExposure_{i,p,1915-18} + Pref_p + X_i + \epsilon_i. \quad (1)$$

Here, CCP_i is a dummy variable indicating whether individual i became a CCP member. We also examine the timing of joining the CCP in our analysis. $NYExposure_{i,p,1919-22}$ and $NYExposure_{i,p,1915-18}$ represent the availability of the *New Youth* during 1915–1918 and 1919–1922 when individual i

was located in prefecture $Pref_p$. $Pref_p$ is a set of prefecture fixed effects that control for various locality characteristics. X_i includes the 13 variables on personal background. We also include five Whampoa-team fixed effects in some specifications to further account for individual experiences at Whampoa.

To account for possible changes in different prefectures before and after 1919, we can further control for the interactions between prefecture fixed effects and the post-1919 dummy ($Pref_p \times Post1919$). However, this reduces the number of observations with variations to approximately half (358 individuals). Nonetheless, we present these results as additional evidence.

Our analysis begins by examining multiple cohorts within our comparison group, which comprises (1) individuals who lived in a locality before 1919, thus missing the post-1919 *New Youth*, (2) those who were in a locality during 1919–1922 but did not experience the post-1919 *New Youth* due to the magazine's exit from the area, and (3) individuals who reached a locality in 1923, subsequently missing the post-1919 *New Youth* because of the magazine's closure (across all localities). We will conduct cohort-by-cohort comparisons to further reinforce our research design. We refer to cohorts (2) and (3) as "the just-missed" group, denoting individuals who were present in localities where the post-1919 *New Youth* was ever available after 1919 but missed it.

In addition to prefecture-level exposures, we also conduct a study on county-level exposures. We compare individuals who lived in the same county and control for county fixed effects with county-level exposures, providing more precise information at the cost of having smaller cells to compare.

We report standard errors clustered at the home county level in our baseline estimations. Moreover, we present standard errors two-way clustered at the prefecture and the year levels. To address potential issues with the parametric assumptions of error structure, we implement a nonparametric permutation test that randomly assigns exposures, following (Chetty et al., 2009).

Balance Tests and Selection into Whampoa Sample. To validate our research design, we conduct two sets of analyses. The first analysis aims to ascertain if the exposed group and the comparison group are similar in personal characteristics. The second analysis checks the potential concerns regarding whether the post-1919 *New Youth* introduced more and different individuals into Whampoa.

Our identification assumption is that, conditional on the localities an individual had been to, $NYExposure_{p,1919–22}$ is close to random. This assumption is reasonable to consider when an individual's relocation was primarily driven by economic opportunities and life events rather than the presence of the *New Youth*. To confirm this, we conduct balance tests with respect to all personal background variables considered. In Columns (1)–(3) of Table 1, we divide individuals into three

groups: (1) those who could access post-1919 *New Youth* in their county, (2) those who could access post-1919 *New Youth* in their prefecture but not their county, and (3) those who could not access post-1919 *New Youth* in their prefecture. Our analyses on prefecture-level exposures are built by comparing (1) and (2) against (3). As shown in Columns (4)–(5), the characteristics between those exposed to the post-1919 *New Youth* and those who weren't are not significantly different once we control for prefecture fixed effects. This confirms the importance of controlling for locality fixed effects to account for the sorting of individuals. Similarly, our analyses on county-level exposures compare (1) against (2) and (3). As shown in Columns (6)–(7), the characteristics between the exposed and everyone else are not significantly different once we control for county fixed effects. Again, this is not true without county fixed effects and confirms the importance of controlling for locality fixed effects.

Furthermore, we investigate whether exposure to the post-1919 *New Youth* correlates with inclusion in our Whampoa sample. While it is true that the *New Youth* was more likely to be available in bigger cities where there might be more a larger pool for Whampoa cadets,[16] our empirical strategy addresses such concerns by including locality fixed effects. In Table 2, we construct a prefecture-level data by periods (before and after 1919) to assess if the duration of *New Youth* circulation influences the number and characteristics of Whampoa cadets. Our findings reveal that, after accounting for prefecture fixed effects, there is no apparent link between the duration of *New Youth* circulation and the quantity of Whampoa cadets, irrespective of the period. Additionally, when examining Whampoa cadets' characteristics, such as educational background and family group proxies, we find no systematic correlations with *New Youth* circulation either.

4 Results

4.1 Impact of Ideology Exposure on CCP Membership

Main Results. Our analysis shows that exposure to post-1919 *New Youth* has a substantial impact on CCP membership. We present the results for prefecture and county-level exposure in Columns (1)–(6) and Columns (7)–(12) of Table 3, respectively. After controlling for prefecture fixed effects in Column (1), we find that the estimate for β in equation (1) is 0.109, indicating a significant positive effect of post-1919 *New Youth* exposure on CCP membership. By contrast, exposure to pre-1919 *New Youth* had little impact on CCP membership (Column (2)), highlighting the importance of ideological content. Column (3) confirms the previous finding by comparing the impacts of post-

[16] We examine the correlations between local characteristics and *New Youth* circulation in Appendix A.7. We find that *New Youth* predominantly circulated in prefectures characterized by larger urban population, larger population, the presence of treaty ports and railways. This correlation did not change greatly before and after 1919.

and pre-1919 *New Youth*.

We include personal characteristics such as education background, birth year fixed effects, team fixed effects, and the number of CCP references in Column (4) to account for the potential confounding factors. The estimate remains similar in Column (5), where we restrict the sample to individuals in Teams 1–4. In Column (6), we further control for additional personal characteristics, such as family background, demographic information, and behavior traits. The estimated coefficient (0.117) in Column (6) serves as our baseline estimate. Columns (7)–(12) present the results for county-level exposure, where we replace prefecture fixed effects with county fixed effects. The estimate for β increases to 0.172, indicating a larger impact for a more precise exposure measure.

Our estimates are similar if we further control for $Pref_p \times Post1919$, indicating that our finding is robust with regard to considering changes in different prefectures before and after 1919 (Table 4). As previously mentioned, 358 individuals would still have variations after including this set of controls, explaining the increase in standard errors compared with our baseline estimates.

Based on the estimate in Column (12), we calculate the persuasion rate following the approach (DellaVigna and Gentzkow, 2010) and obtain a persuasion rate of 18.9%.[17] The relatively high persuasion rate is likely because we are studying a group of radical youth who were making a sharp choice in party identity, and were therefore more susceptible to persuasion.

Among the personal characteristics, having a personal CCP reference in most of the specifications is associated with an increased chance of joining the CCP. This is consistent with research on the importance of social networks in political participation. In contrast, socioeconomic variables are not predictive of CCP membership. Our examination of the interaction effects between personal characteristics and post-1919 *New Youth* exposure (Appendix B.1) does not reveal strong heterogeneities, except for some complementarity between ideology exposure and social networks, as well as between ideology exposure and having a business family background. The relatively insignificant impact of socioeconmic characteristics may be due to the high risk the CCP faced during this period, making it difficult to identify specific characteristics that may have been more advantageous for joining the revolution.

The impact of post-1919 *New Youth* exposure is robust to various ways of calculating standard errors (Appendix B.2). Moreover, a nonparametric permutation test by randomly assigning the treated sample 10,000 times confirms the significance of our estimate on post-1919 *New Youth* (Figure 5). Since these results do not rely on parametric assumptions, they further support the relevance of post-1919 *New Youth* exposure. Additionally, we also employ a continuous measure for $NYExposure_{p,1919\text{–}22}$ by counting the share of years exposed and find a pattern similar to using

[17]This is a lower bound because we assume the reading rate in treated counties to be 1 in this calculation. Our analysis across heterogenous localities below provides estimates to shed light on possible influence on reading rate.

a dummy variable (Appendix B.3). For ease of interpretation, we focus on the dummy variable of exposure in the following analyses.

A Closer Look at Possible Endogeneity Concerns. While we have accounted for time-invariant locality characteristics and changes in their influence before and after 1919 in our main analyses, we acknowledge potential endogeneity concerns and address them in this section. Specifically, we examine the endogenous entry of the *New Youth* and possible relocation preferences of those interested in reading it.

First, we find that the impact of $NYExposure_{p,1915-18}$ on CCP membership is insignificant, suggesting that endogenous entry of the *New Youth* is not a critical concern (Table 3). Furthermore, when we separate localities that always had access to the magazine from those that experienced its entry and exit, the estimates are similar, as shown in Table 5. These results suggest that the endogenous entry of the *New Youth* is not a critical concern.

To address concerns about relocation preferences, we construct placebo relocations by assuming an individual's location was in the previous or next locality in their life history. We use these placebo relocations to calculate placebo exposures to post-1919 *New Youth*. If relocation preferences were driving our results, we would expect placebo exposures to explain the impact of $NYExposure_{p,1919-22}$. However, Table 6 shows that placebo exposures cannot explain our finding, suggesting that our main result is unlikely to be driven by individuals' relocation preferences.

Finally, as mentioned above, our comparison groups comprises of (1) individuals who lived in a locality before 1919, thus missing the post-1919 *New Youth*, (2) those who were in a locality during 1919–1922 but did not experience the post-1919 *New Youth* due to the magazine's local exit, and (3) individuals who reached a locality in 1923, subsequently missing the post-1919 *New Youth* because of the magazine's national closure. If our main finding were confounded by certain changes in individuals' sorting over time, we would anticipate different coefficients when using different comparison groups. However, as shown in Figure 6 and Appendix Table B.4, when comparing with these different subgroups, our estimates are similar to our baseline one, suggesting that sorting of individuals is unlikely to be the driver of our main finding.

Measurement Error. Measurement error is likely in our setting. In Appendix B.5, we conduct three sets of analyses at the locality and individual levels to examine the robustness of our results. First, we test the sensitivity of our estimate by dropping major metropolitan cities (Beijing, Shanghai, Guangzhou). Our finding remains robust, indicating that our results are not driven by a few large cities. In our subsequent analysis of the persuasion rate, we further examine the heterogeneity across localities.

Second, we investigate the precision of life events in an individual's records. We find that our results are not sensitive to the number of life events recorded, as our findings hold even when we drop those with few or many life events. Thus, the abundance or lack of records on individuals does not affect our results.

Third, we infer certain missing years based on an individual's other records with concrete year information and formulate a calculation against the timeline of the education system in this era. To check whether our method is reasonable, we apply the same inference method by pretending that we only knew the birth year, end year, and life events for all individuals. As shown in Appendix B.5, the estimate is similar to our baseline if we apply the same inference method to all our sample, which solidifies the validity of inferring missing years on an individual's records.

Heterogeneous Circulation Locations. Our research design delivers an intention-to-treat estimate without using information on reading rate. To gauge the influence of reading rate, we conduct three sets of analysis across localities. First, we separate the prefecture-level post-1919 *New Youth* into a breakdown of within a county and outside the county. Columns (1)–(2) of Table B.6 show little impact if the post-1919 *New Youth* was available outside of a cadet's located county, implying that the reading rate is small if the magazine was not available within the same county.

Second, we separate the circulation county into small and big cities, using an urban population size threshold of 100,000. Using the same specification as in our baseline, we obtain estimates of 0.26 for small cities and 0.13 for big cities (Columns (3)–(4) of Appendix B.6). One possibility is that the reading rate was higher if the *New Youth* was available within a small city.

Third, a subsample of the circulation points were bookstores located in schools. Based on the life history of the cadets, we know whether a cadet was studying in those schools. If the *New Youth* was available at a cadet's school, it seems reasonable to assume the reading rate being close to 1 given that the *New Youth* was the most discussed periodical in this era. When separating this very local exposure from others, we find its coefficient to be 0.26 whereas that of the rest is 0.15 (Columns (5)–(6)), which implies a persuasion rate of 29.2%. These patterns support that within our studied individuals, our baseline persuasion rate is likely to be a lower bound.

Potential Spillover Effects. Although we cannot directly observe how the cadets influenced each other, we can use proxies of social connections to check for potential spillover effects among cadets. Specifically, we examine whether individuals are more likely to join the CCP if their connected group has a higher degree of ideology exposure. We define one's connected group in two ways: the first refers to their hometown group, and the second concerns their social group connected via shared references. As reported in Appendix B.7, we do not find strong evidence of spillover from the hometown group, but we do find suggestive evidence of spillover from the group connected via

shared references. However, considering these potential spillover effects does not alter our main finding.

Additional Results from Cohorts 2–4. We also conduct an analysis on the 700 cadets from the second to fourth cohorts whose biographical information is available. As shown in Appendix B.8, we find that ideology exposure also matters for these cohorts, who were enrolled during 1925 and 1926. Therefore, our finding on the first cohort appears to apply to other cohorts whose information is less systematically available.

4.2 Interpreting Post-1919 *New Youth* Exposure

Our use of post-1919 *New Youth* exposure as a measure of Communist ideology exposure is motivated by historical narratives (Section 2.1). However, similar to those who left personal accounts on their reading experience, the cadets likely read other periodicals and were exposed to various activist organizations during the New Culture Movement, which may have also contributed to the recruitment of the CCP. In this section, we examine whether post-1919 *New Youth* was reasonable proxy for Communist ideology diffusion by leveraging archival data on pro-Communism periodicals and activist groups in this era, the timing of joining the CCP, and proxies for potential social interactions.

The Network of Pro-Communism Periodicals. The *New Youth* inspired other pro-Communism periodicals to start, with the editors of the *New Youth* helping to co-found several of those periodicals. Based on a series of archives published as *Red Collections: Progressive Periodicals from 1915–1949* (Xue, 2014), we gather information on the editors and editorial offices for all 19 pro-Communism periodicals that existed during 1919–1922 in these archives. We define a link between two periodicals as sharing an editor or editorial office.

We plot the network among these pro-Communism periodicals in Figure 7. As shown by the closeness centrality index noted in the figure, the *New Youth* is indeed the most central periodical in the network. In comparison to the *New Youth*, the other periodicals started later and thus had a shorter period of influence during 1919–1922.

We also gather information on the circulation locations of two other pro-Communism periodicals in the network, *Young China* and the *Guide Weekly*. We specifically chose these because of their closeness in circumstance to the *New Youth*. *Young China* had the same length of existence as the post-1919 *New Youth* during 1919–1922 and the *Guide Weekly* is the second central periodical in the network. We find that exposures to these additional periodicals can be considered a subset of exposure to the post-1919 *New Youth*: 47% of our studied cadets were exposed to the circulation of

these two periodicals, out of which 99.4% overlapped with post-1919 *New Youth* exposure. When separating individuals out between those exposed to post-1919 *New Youth* only and those exposed to all three periodicals, we find similar estimates (Panel (a) of Table 7), supporting the importance of post-1919 *New Youth* as a reasonable proxy for Communism diffusion.

Activist Groups and Communist Organizations. There existed various voluntary groups in the New Cultural Movement. We focus on the 28 of them that have been characterized as left-wing activist groups from archival information of their reports and meeting minutes (Zhang et al., 1979, Xu, 2013).[18] To check the relationship between these organizations and ideology exposure, we code an exposure measure to these organizations using the same way we code exposure to the *New Youth*. When examining the relationship between these exposures, we find a heavy overlapping pattern: 31% of our studied individuals were exposed to those activist groups, out of which 99.1% overlapped with post-1919 *New Youth* exposure. When separating out the individuals that were only exposed to post-1919 *New Youth* and those exposed to both post-1919 *New Youth* and these 28 activist groups, we find an estimate of 0.11 for the former and an estimate of 0.18 for the latter (Panel (b) of Table 7). Thus, exposure to post-1919 *New Youth* has a major impact in localities even without these activist groups.

Further Evidence. Despite all the evidence, one may still wonder whether the post-1919 *New Youth* captures non-idea factors such as CCP recruitment and non-Communism social interactions in a locality. To check the importance of such concerns, we examining the timing of joining the CCP and the relevance of potential social interactions for our finding.

If post-1919 *New Youth* captured local recruitment effort, we would expect individuals to join the CCP in the locality where they were exposed (i.e., before moving to Whampoa). However, as discussed above, few could access Communist organizations and only a small portion of CCP members in our sample joined before the Whampoa period (i.e., during 1921–1924).[19] We analyze the effect of post-1919 *New Youth* exposure by the year of CCP membership and find that pre-Whampoa CCP memberships can only explain one third of our main finding (Appendix B.9). Additionally, using references to the CCP to proxy for the party's stronghold in a certain locality, we find a strong influence of having CCP references for those who joined the CCP before the Whampoa period, but not for those who joined during 1924–1928. This suggests that our main finding regarding the role of ideas is unlikely to be explained by local recruitment efforts alone.

Moreover, for each cadet, we know who co-located with him before joining the Whampoa and

[18] Zhang et al. (1979) provide the data source. Xu (2013) studies why five of these 28 groups became Communist groups. We follow Xu (2013) in identifying these groups.

[19] However, none of the cadets in our sample joined the CCP before getting exposed to post-1919 *New Youth*.

can calculate the average post-1919 *New Youth* exposure ratio among his co-locaters at Whampoa. If post-1919 *New Youth* captured primarily social interactions rather than ideas, we would expect a strong heterogoneous pattern: After arriving at Whampoa, those who were more likely to share experience with his exposed co-locaters would be more likely to join the CCP. In the data, however, we do not find any strong heterogeneity in the impact of idea exposure with respect to such potential social interactions (Appendix B.10), which again shows the difficulty to attribute our main finding to the role of non-idea factors alone.

In sum, we interpret the role of post-1919 *New Youth* as a reasonable proxy for Communism diffusion. Although a subgroup of our studied individuals could access additional pro-Communism periodicals and organizations, considering the additional sources does not change our main finding.

4.3 Subsequent Political Outcomes

Can individuals mobilized by ideas behave differently in critical historical moments? We aim to shed light on this important question by exploring information on the subsequent careers of the Whampoa cadets. During the 1920s–1940s, the CCP and the KMT experienced various political struggles, resulting in many members quitting their parties and even sacrificing their lives. For instance, in 1927, the KMT under Chiang Kai-shek's direction conducted a purge to eliminate Communist influence, causing the CCP to lose over 90% of its members. The two civil wars (1927–1936 and 1945–1949) and the anti-Japanese war (1937–1945) were also significant conflicts that caused both parties to experience membership losses.

We use two methods to examine individual outcomes in subsequent political conflicts. Our first method compares four groups of individuals: CCP members with and without post-1919 *New Youth* exposure, and KMT members with and without post-1919 *New Youth* exposure.[20] Table 8(a) presents our findings, using KMT members without post-1919 *New Youth* exposure as the reference group. We use seemingly unrelated regressions to allow for correlated error terms across the equations (for quitting and sacrifice). Our results show that CCP members are more likely to quit or sacrifice than KMT members, reflecting greater uncertainty faced by the CCP. Furthermore, among CCP members, those who were exposed to post-1919 *New Youth* were 23 percentage points less likely to quit the party and 25 percentage points more likely to sacrifice their lives (Columns (1)–(2)).

Upon closer inspection, we find that our main finding is primarily driven by the period from 1927 to 1936, during which the CCP was weaker and faced severe attacks from the KMT. Out of

[20]Although all cadets was initially a KMT, they had to make a choice between the KMT and the CCP after 1927. Our KMT sample refers to those who did not join the CCP.

the 66 CCP members who sacrificed their lives, 27 died during 1924–27, and 33 died during the first KMT-CCP Civil War. Out of the 38 CCP members who quit the party, 5 quit during 1924–27, and 33 quit during the first KMT-CCP civil war. A similar pattern emerges when we zoom into the period of the first civil war (Columns (3)–(4)).

Our second approach employs a multinomial probit model to analyze four outcomes: (1) not being a member of the CCP, (2) having been a CCP member but later quitting, (3) being a CCP member and surviving, and (4) being a CCP member and having sacrificed.[21] As indicated in Column (2) of Table 8(b), our findings align with those from our initial method. Specifically, exposure to the post-1919 *New Youth* shows a positive association with the fourth outcome (sacrifice their lives) comparing with the second outcome (quitting the CCP).

To better comprehend the extent to which our findings are mechanically driven by political assignments, we further examine both the leadership positions held by cadets in Whampoa Academy and their subsequent army assignments after graduation (see Appendix B.11). We find no significant association between post-1919 *New Youth* exposure and these assignments and controlling for these assignments does not vary our finding on sacrifice or quitting. These results strengthen the argument that individuals mobilized by ideas have the capacity to exhibit distinct behaviors and make diverse choices, particularly in times of turmoil. However, it is important to acknowledge that pivotal individual decisions during historical junctures are influenced by a myriad of factors. As highlighted by historians, the majority of revolutionaries underwent a complex process of transformation, gradually becoming committed Communists (Van de Ven, 2023, Esherick, 1995). Therefore, our findings do not imply predetermined future political outcomes but rather suggest that ideology exposure can interact with other elements to shape personal choices.

5 Conclusion

Ideas and their diffusion may have played pivotal roles in many social and political movements. One may even claim: "no ideas, no revolution" (Brinton, 1938). However, quantifying the influence of ideas presents significant challenges. Our study focuses on the impact of Communist ideology, a subject that, despite its global significance, has seen limited exploration by social scientists. By merging data on the dissemination of Communist ideas with detailed biographical histories of key individuals, we have shed light on how exposure to these ideas shaped the political choices of future leaders during pivotal moments in China's modern history.

It is essential to recognize that ideological exposure is just one of many factors influencing

[21] We choose multinomial probit over multinomial logit in our setting, as the former does not require the strict independence of irrelevant alternatives assumption across the four outcomes.

political choices. In the context of Communism's diffusion in China, several macro-level factors played crucial roles, including the implications of the Treaty of Versailles, Soviet and Comintern strategies, and the prevailing weakness of the Chinese state. These elements form the backdrop of our study. On an individual level, we find that, apart from ideological influence, the political identities of personal connections also significantly predict an individual's political leaning, highlighting the potent impact of social networks. Conversely, the correlation between economic background and political identity in our study's context was minimal. The particular relevance of ideology, in contrast with economic status, provides new evidence to the rich empirical literature on conflict participation that often focuses on various forms of interests. We hypothesize that ideas may be especially pertinent in understanding leaders, particularly those who join movements at their inception, a notion that remains to be tested in other settings.

Understanding individual participation in high-risk social and political movements is complex. Our study offers a possibility to understand the political decisions of individuals by examining their personal profiles and choices. This approach could be applicable in other contexts where detailed information on leadership figures is often readily available, offering new research opportunities in diverse areas.

References

Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhuravskaya (2015). Radio and the rise of the nazis in prewar germany. *The Quarterly Journal of Economics 130*(4), 1885–1939.

Ang, Desmond (2023). The birth of a nation: Media and racial hate. *American Economic Review 113*(6), 1424–1460.

Ash, Elliott, Daniel L Chen, and Suresh Naidu (2022). Ideas have consequences: The impact of law and economics on american justice. Technical report, National Bureau of Economic Research.

Bai, Ying, Ruixue Jia, and Jiaojiao Yang (2023). Web of power: How elite networks shaped war and politics in china. *The Quarterly Journal of Economics 138*(2), 1067–1108.

Bailyn, Bernard (1967). *The ideological origins of the American Revolution*. Harvard University Press.

Blattman, Christopher and Edward Miguel (2010). Civil war. *Journal of Economic literature 48*(1), 3–57.

Brinton, Crane (1938). *The Anatomy of Revolution*. New York: Vintage Books. Reprinted in 1965.

Cagé, Julia, Anna Dagorret, Pauline Grosjean, and Saumitra Jha (2023). Heroes and villains: The effects of heroism on autocratic values and nazi collaboration in france. *American Economic Review 113*(7), 1888–1932.

Cantoni, Davide, Yuyu Chen, David Y Yang, Noam Yuchtman, and Y Jane Zhang (2017). Curriculum and ideology. *Journal of political economy 125*(2), 338–392.

Cantoni, Davide, Andrew Kao, David Y Yang, and Noam Yuchtman (2023). Protests. Technical report, National Bureau of Economic Research.

Central Executive Committee of the Kuomintang (1924). *Lujun junguan xuexiao xiangxi diaochabiao[Detailed Survey for the First to Fourth Teams of the First Cohort of the Military Academy]*. Taipei: Wenhai Press Reprinted.

Chen, Duxiu (1916). Wuren zuihou zhi juewu. *Qingnian zazhi 1*.

Chen, Ting and James Kai-Sing Kung (2020). The rise of communism in china. *NBER Working Paper*.

Chen, Yuhuan (2017a). *Whampoa Junxiao Diyiqisheng Quanjilu [Complete Records on the First-Cohort Cadets of the Whampoa Military School]*. Taipei: Sixing Wenhua Press.

Chen, Yuhuan (2017b). *Whampoa Junxiao Dierqisheng Quanjilu [Complete Records on the Second-Cohort Cadets of the Whampoa Military School]*. Taipei: Sixing Wenhua Press.

Chen, Yuhuan (2017c). *Whampoa Junxiao Disanqisheng Quanjilu [Complete Records on the Third-Cohort Cadets of the Whampoa Military School]*. Taipei: Sixing Wenhua Press.

Chen, Yuhuan (2018). *Whampoa Junxiao Disiqisheng Quanjilu [Complete Records on the Fourth-Cohort Cadets of the Whampoa Military School]*. Taipei: Sixing Wenhua Press.

Chetty, Raj, Adam Looney, and Kory Kroft (2009). Salience and taxation: Theory and evidence. *American economic review 99*(4), 1145–1177.

Chow, Tse-tsung (1960). *The May fourth movement: Intellectual revolution in modern China*. Harvard University Press.

Collier, Paul and Anke Hoeffler (2004). Greed and grievance in civil war. *Oxford economic papers 56*(4), 563–595.

DellaVigna, Stefano and Matthew Gentzkow (2010). Persuasion: empirical evidence. *Annu. Rev. Econ. 2*(1), 643–669.

DellaVigna, Stefano and Ethan Kaplan (2007). The fox news effect: Media bias and voting. *The Quarterly Journal of Economics 122*(3), 1187–1234.

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

Dirlik, Arif (1989). *The origins of Chinese communism*. Oxford University Press.

Esherick, Joseph W (1995). Ten theses on the chinese revolution. *Modern China 21*(1), 45–76.

Fan, Guorui and Thomas S Popkewitz (2020). *Handbook of Education Policy Studies*. Springer.

Fearon, James D and David D Laitin (2003). Ethnicity, insurgency, and civil war. *American political science review 97*(1), 75–90.

Graff, David Andrew and Robin Higham (2012). *A military history of China*. University Press of Kentucky.

Jha, Saumitra and Steven Wilkinson (2023). Revolutionary contagion.

Johnson, Chalmers A (1962). *Peasant nationalism and communist power: the emergence of revolutionary China, 1937-1945*. Stanford University Press.

Jones, Benjamin F and Benjamin A Olken (2005). Do leaders matter? national leadership and growth since world war ii. *The Quarterly Journal of Economics 120*(3), 835–864.

Magness, Phillip W and Michael Makovi (2023). The mainstreaming of marx: Measuring the effect of the russian revolution on karl marx's influence. *Journal of Political Economy 131*(6), 000–000.

Meisner, Maurice (1999). *Mao's China and after: A history of the People's Republic*. Simon and Schuster.

Mokyr, Joel (2005). The intellectual origins of modern economic growth. *The Journal of Economic History 65*(2), 285–351.

Ottinger, Sebastian and Lukas Rosenberger (2023). The american origin of the french revolution.

Prat, Andrea and David Strömberg (2013). The political economy of mass media. *Advances in economics and econometrics 2*, 135.

Schwarcz, Vera (1986). *The Chinese enlightenment: Intellectuals and the legacy of the May Fourth Movement of 1919*,

Volume 27. Univ of California Press.

Shadmehr, Mehdi (2017). Khomeini's theory of islamic state and the making of the iranian revolution. *Available at SSRN 3037793*.

Smith, Stephen Anthony (2014). *The Oxford handbook of the history of communism*. OUP Oxford.

Snow, Edgar (1937). *Red Star over China: The Classic Account of the Birth of Chinese Communism*. Victor Gollancz Ltd.

Spenkuch, Jörg L, Edoardo Teso, and Guo Xu (2023). Ideology and performance in public organizations. *Econometrica 91*(4), 1171–1203.

Taylor, Jay (2009). *The generalissimo: Chiang Kai-shek and the struggle for modern China*. Harvard University Press.

Van de Ven, Hans J (2023). *From friend to comrade: the founding of the Chinese Communist Party, 1920-1927*. Univ of California Press.

Xu, Xiaohong (2013). Belonging before believing: Group ethos and bloc recruitment in the making of chinese communism. *American Sociological Review 78*(5), 773–796.

Xue, Muqiao (Ed.) (2014). *Hongcang: Jinbu Qikan Zonghui (1915–1949) [Red Collections Progressive Periodicals from 1915–1949]*. Xiangtan: Xiangtan University Press.

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

Yang, Kuisong (1994). Mosike xinfaxian de deng xiaoping zizhuan ["autobiography of deng xiaoping" newly discovered in moscow]. *Ming Pao Monthly 9*.

Yeh, Wen-hsin (2023). *Provincial passages: culture, space, and the origins of Chinese communism*. Univ of California Press.

Yoshihiro, Ishikawa (2012). *The Formation of the Chinese Communist Party*. Columbia University Press.

Yun, Daiying (1981). *Yun Daiying riji [Diaries of Yun Daiying]*. Beijing: Central Party School Press.

Zhang, Guotao (1971). *Wo de huiyi [My Reflections]*. Hong Kong: Ming Pao Monthly Press.

Zhang, Yunhou, Xuli Yin, Qingxiang Hong, and Yunkai Wang (Eds.) (1979). *Wusi shiqi de shetuan [Voluntary Societies during the May Fourth Era]*. Beijing: Sanlian chubanshe.

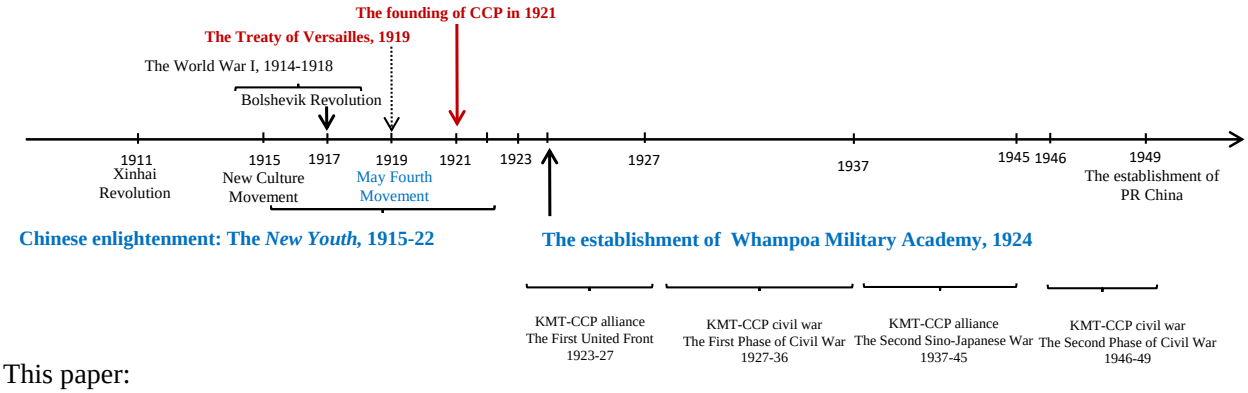
Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov (2020). Political effects of the internet and social media. *Annual review of economics 12*, 415–438.

Figure 1: Ideological Transition of the *New Youth*

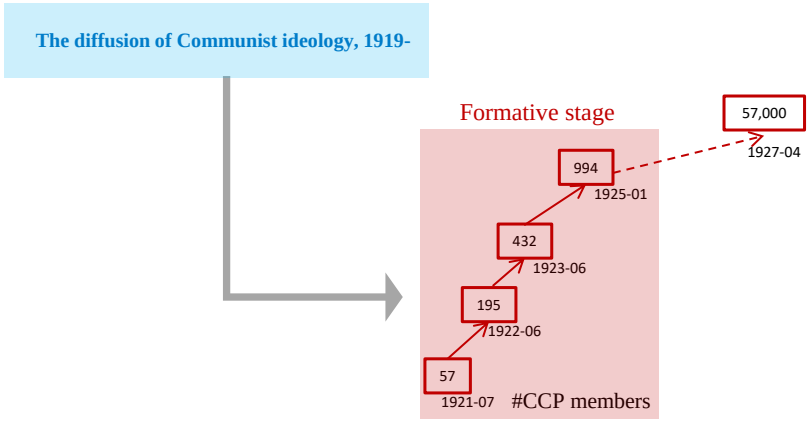
Note. This figure shows the content change of the *New Youth* around 1919–20. In Panel (b), the "Communism" topic includes the phrases "Marx", "Lenin", "socialism", "Communism", "Soviet", "Bolshevism", "revolution", "labor", "worker", "strike", and "proletariat". "Literature" includes all novels, dramas, literary critics, essays on linguistics and philology, and book-reading notes. We exclude poems when calculating these shares.

Figure 2: Key Events (1911–1949)

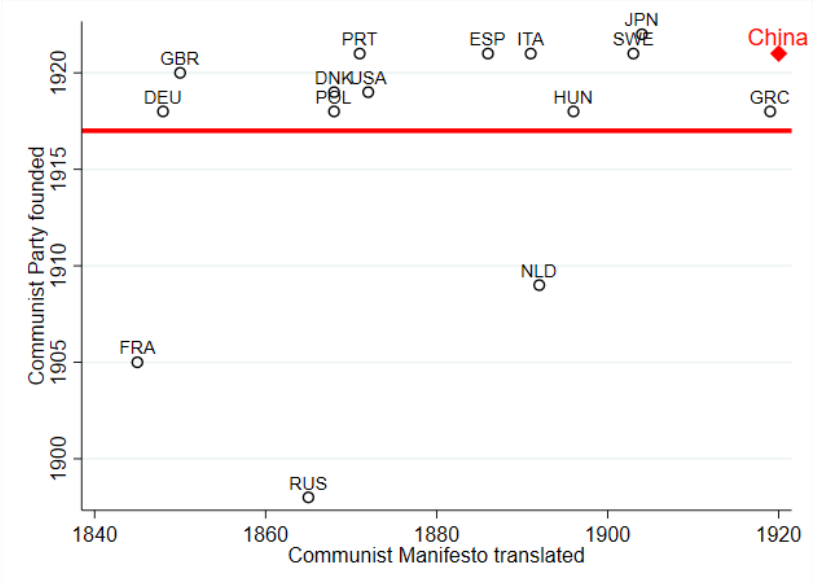
(a) Timeline



This paper:



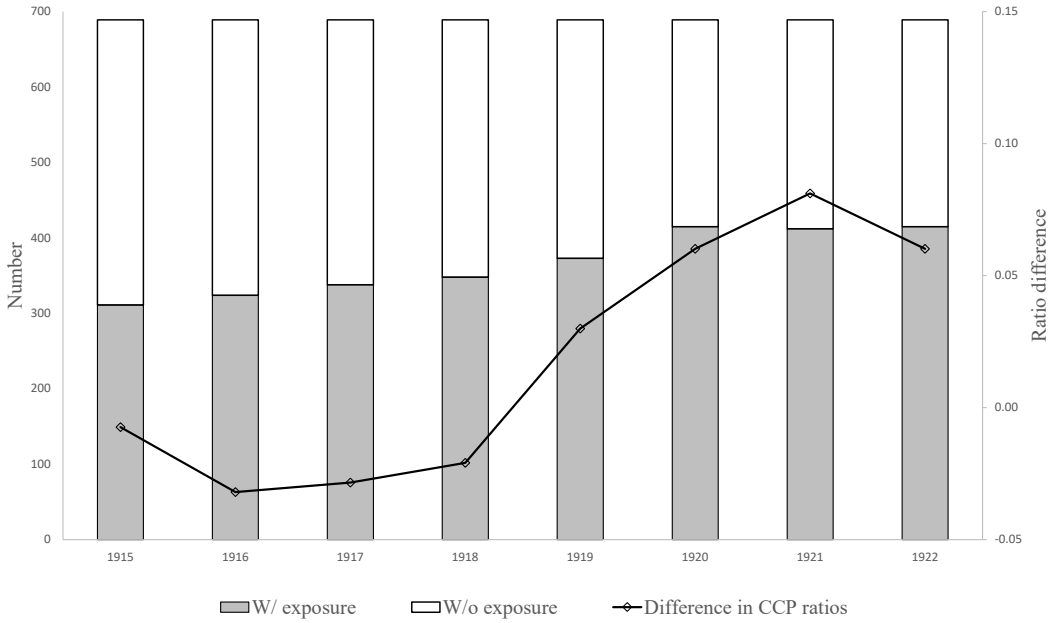
(b) Communist Parties vs. the Communist Manifesto



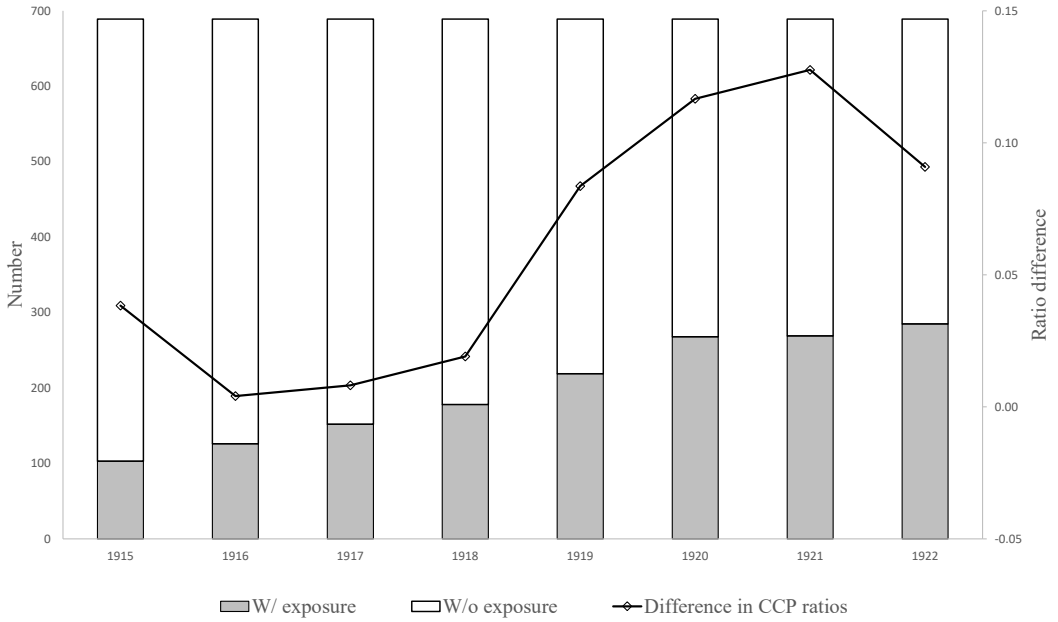
Note. Panel (a) presents the major events in China during 1911–1949. Our analyses focus on how exposure to the *New Youth* (1919–22) affected the Whampoa cadets' choice to become a CCP member in CCP's formative stage in the 1920s. Panel (b) plots the founding years of Communist parties and the years of Communist Manifesto translation across countries.

Figure 3: Raw Data I: Exposure to the *New Youth* and CCP Membership

(a) Prefecture-level



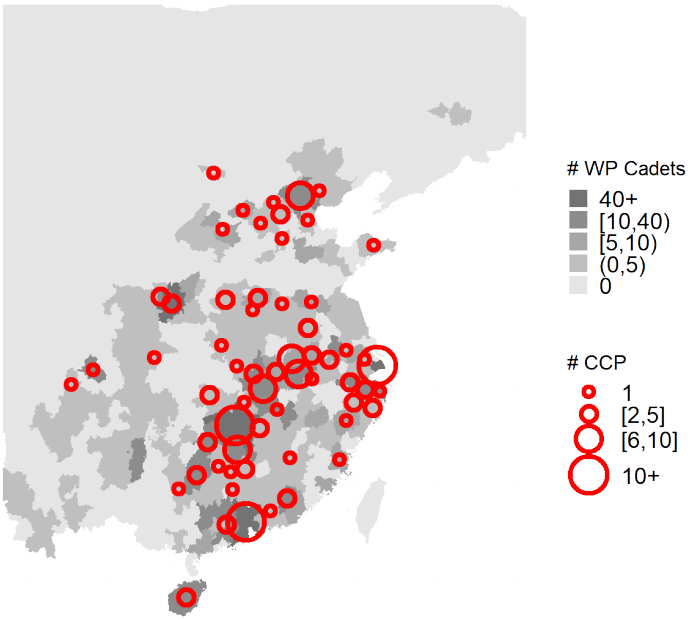
(b) County-level



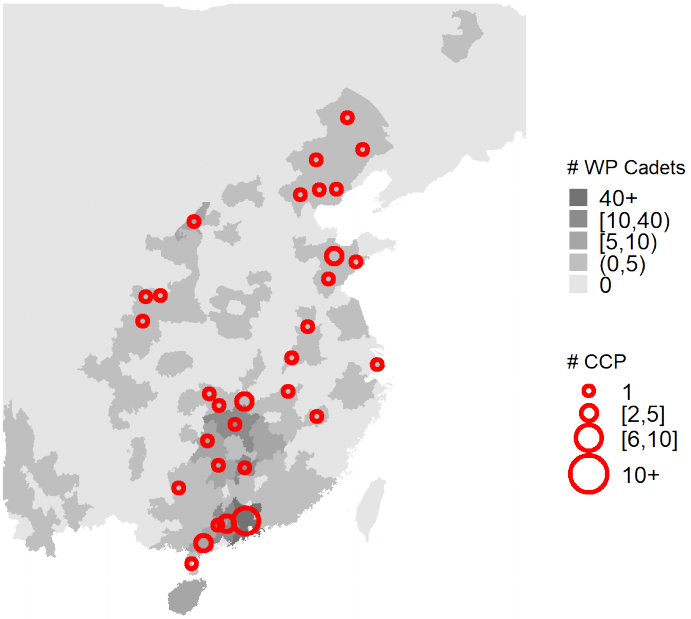
Note. The shaded and unshaded areas in the bar represent the number of cadets with and without *New Youth* exposure in each year respectively. The solid line with diamond markers represents the difference in CCP ratios between the cadets with *New Youth* exposure and those without. As shown, the differences became positive after 1919.

Figure 4: Raw Data II: Post-19 *New Youth* Exposure and CCP Membership

(a) CCP w/ Post-19 *New Youth* Exposure



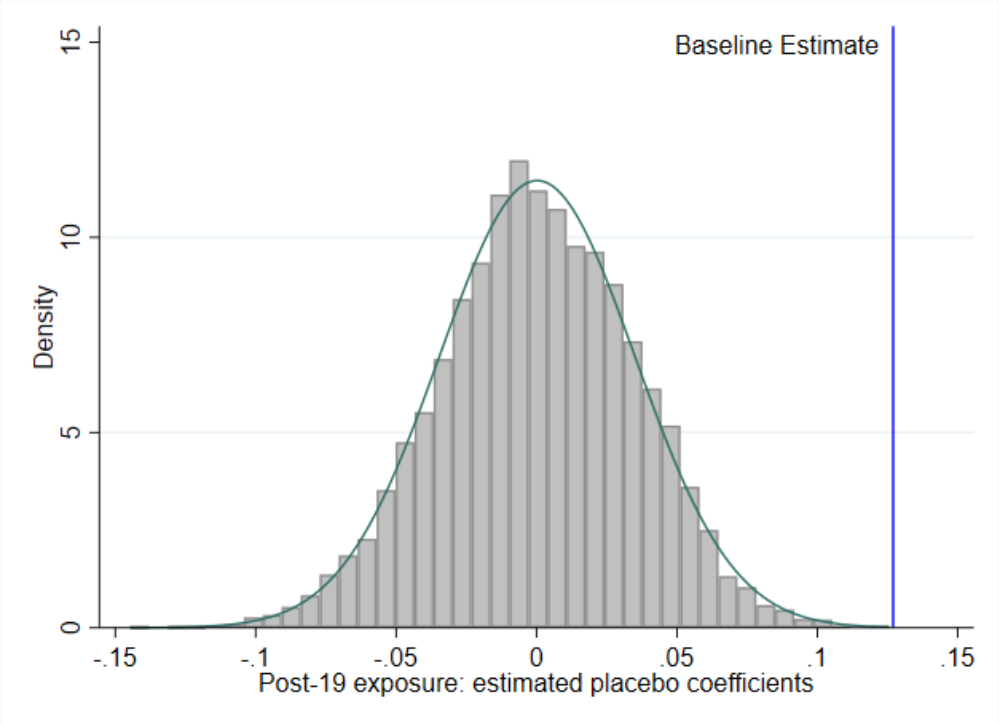
(b) CCP w/o Post-19 *New Youth* Exposure



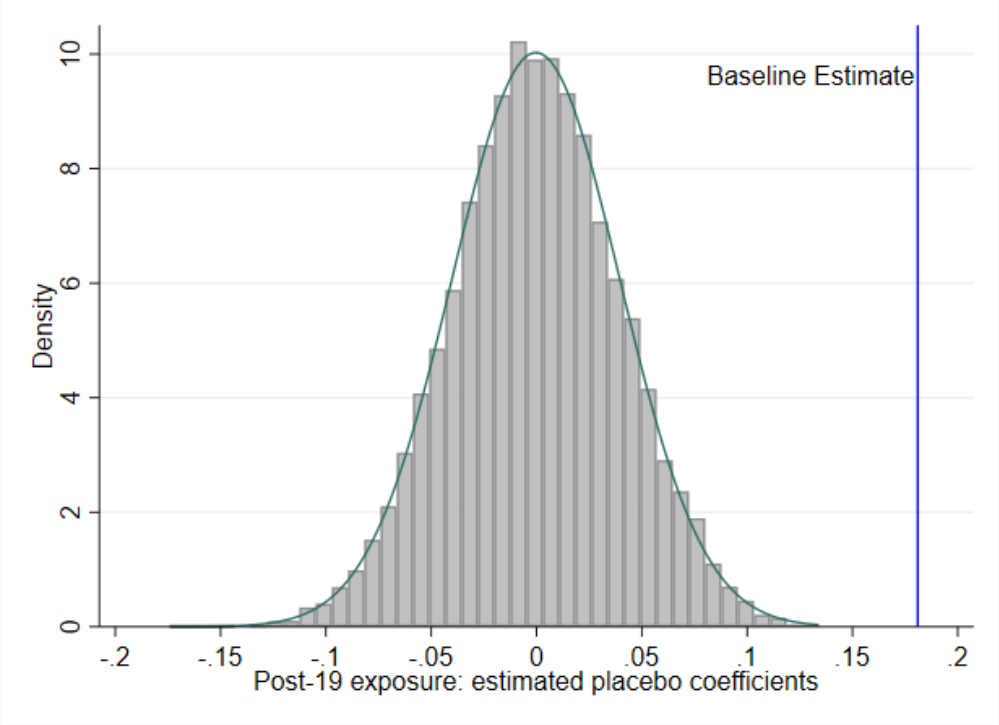
Note. These figures plot the number of CCP members and the number of cadets across the groups with or without post-19 *New Youth exposure*.

Figure 5: Distribution of Placebo Estimates

(a) Exposure Defined at Prefecture Level

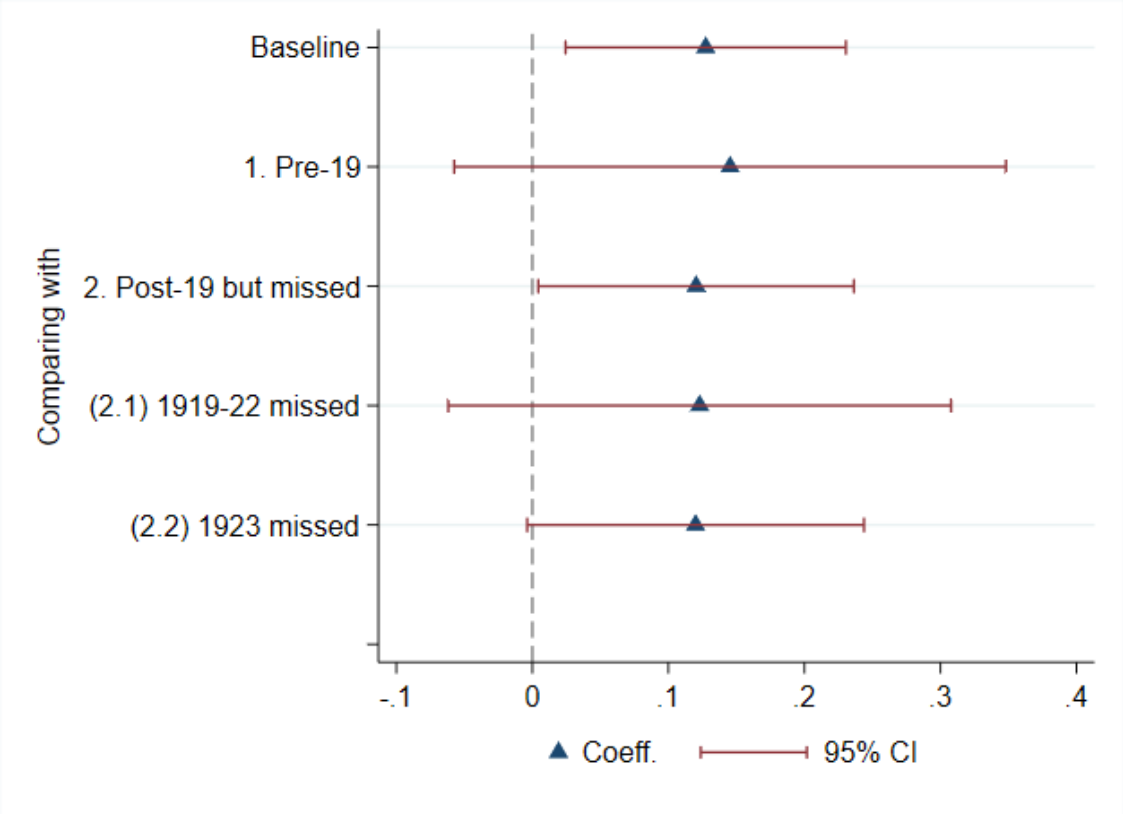


(b) Exposure Defined at County Level



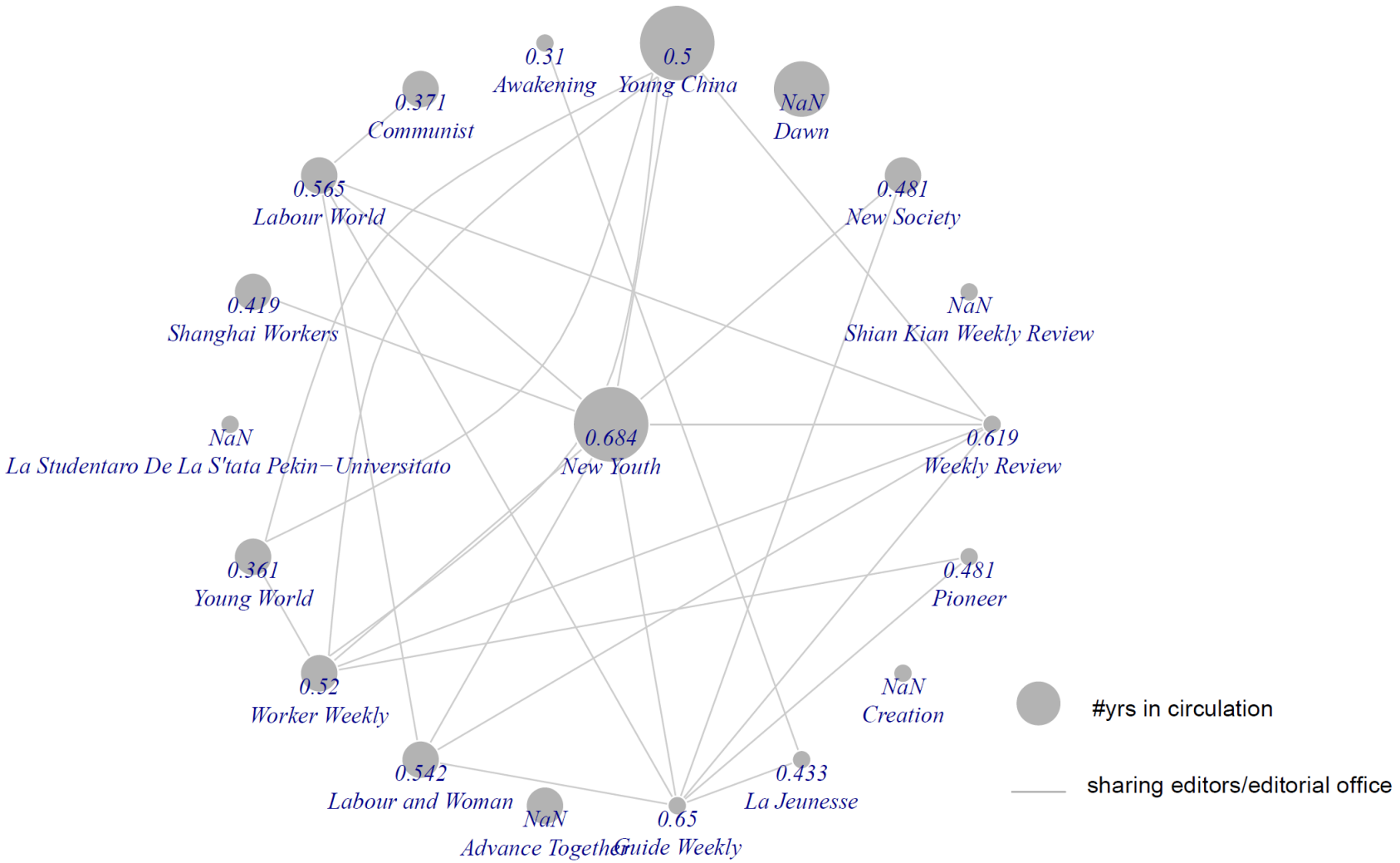
Note. These figures plot estimates from randomly assigned exposures. They demonstrate that our estimates are significantly different from these randomly assigned treatment effects.

Figure 6: Comparison Groups by Cohorts



Note. This figure plots the estimates when comparing with different cohorts that had resided in prefectures ever with post-1919 *New Youth*. These estimates show a similar pattern to our baseline results.

Figure 7: The Network of Pro-Communism Periodicals



Note. This figure plots links among all 19 Pro-Communism periodicals in existence during 1919–1922. A link means that two periodicals share editors or an editorial office. The size of the circles indicates the number of circulation years during 1919–1922, and the number in the circle indicates the closeness centrality of a periodical in this network.

Table 1: Balance Tests

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	New Youth exposure, 1919-22			Differences			
Prefecture-level exposure	=1		=0	(1)+(2) vs. (3)		(1) vs. (2)+(3)	
County-level exposure	=1	=0		w/o Pref. FE	**w/ Pref. FE**	w/o county FE	**w/ county FE**
Sample: All							
Personal characteristics I, N=	365	143	181				
Birth year	1900.5	1901.2	1900.7	0.009	0.297	-0.445**	0.223
	(2.889)	(2.802)	(2.911)	[0.260]	[0.457]	[0.218]	[0.491]
Education: Middle school and above	0.981	0.930	0.928	0.038*	0.012	0.052***	0.001
	(0.137)	(0.256)	(0.259)	[0.023]	[0.037]	[0.016]	[0.042]
Education: vocational	0.633	0.615	0.541	0.087	-0.055	0.059	-0.020
	(0.483)	(0.488)	(0.500)	[0.055]	[0.065]	[0.051]	[0.108]
Education: college	0.164	0.035	0.050	0.078***	0.000	0.121***	-0.032
	(0.371)	(0.184)	(0.218)	[0.025]	[0.035]	[0.024]	[0.050]
Networks: # CCP references	0.263	0.049	0.116	0.087*	0.069	0.177***	0.090
	(0.608)	(0.247)	(0.570)	[0.051]	[0.084]	[0.042]	[0.113]
Sample: Teams 1-4:							
Personal characteristics II, N=	314	85	131				
Family: gentry	0.280	0.188	0.214	0.047	0.042	0.077**	0.132
	(0.450)	(0.393)	(0.412)	[0.043]	[0.082]	[0.039]	[0.118]
Family: business	0.303	0.294	0.328	-0.027	0.007	-0.012	-0.058
	(0.460)	(0.458)	(0.471)	[0.049]	[0.082]	[0.041]	[0.139]
Family economic status (1-3)	1.675	1.518	1.611	0.031	0.102	0.101**	0.051
	(0.521)	(0.569)	(0.549)	[0.060]	[0.098]	[0.050]	[0.148]
# Siblings	3.274	3.624	3.489	-0.140	0.182	-0.268	-0.292
	(2.258)	(2.502)	(2.761)	[0.254]	[0.558]	[0.216]	[0.879]
Married	0.506	0.471	0.679	-0.181***	-0.141	-0.091*	0.005
	(0.501)	(0.502)	(0.469)	[0.051]	[0.092]	[0.048]	[0.146]
Having children	0.146	0.129	0.122	0.021	0.001	0.021	0.073
	(0.354)	(0.338)	(0.329)	[0.035]	[0.070]	[0.031]	[0.102]
Religious	0.099	0.129	0.122	-0.017	0.045	-0.026	-0.006
	(0.299)	(0.338)	(0.329)	[0.035]	[0.047]	[0.029]	[0.073]
Smoking and drinking	0.035	0.106	0.053	-0.003	0.028	-0.039*	-0.100
	(0.184)	(0.310)	(0.226)	[0.021]	[0.032]	[0.022]	[0.066]

Note. Standard deviations are reported in the parentheses in Columns (1)–(3). Standard errors clustered at the home county level are presented in brackets in Columns (4)–(7). Once locality fixed effects are included, the groups with or without exposure are balanced across all characteristics (Columns (5) and (7)).

Table 2: *New Youth* Circulation and Cadets in Our Sample

Dependent variable:	# Whampoa cadets					
			some college	family: gentry	family: business	family: middle/rich
	(1)	(2)	(3)	(4)	(5)	(6)
(1): # years with New Youth	0.285	0.289	-0.057	0.072	-0.018	0.122
	(0.284)	(0.283)	(0.085)	(0.068)	(0.117)	(0.186)
(2): # years with New Youth × 1919-22		-0.033	0.152	-0.021	-0.068	0.021
		(0.175)	(0.092)	(0.073)	(0.072)	(0.129)
Controls × Period	Y	Y	Y	Y	Y	Y
Prefecture FE	Y	Y	Y	Y	Y	Y
Period FE	Y	Y	Y	Y	Y	Y
Observations	512	512	512	512	512	512
R-squared	0.619	0.619	0.212	0.582	0.593	0.601
#Prefectures	256	256	256	256	256	256
(1)+(2): # years with New Youth, 1919-22		0.256	0.095	0.052	-0.086	0.143
		(0.335)	(0.111)	(0.096)	(0.131)	(0.219)

Note. These estimates are based on a prefecture-level panel dataset across two periods (before and after 1919). Controls include city population, total population, the presence of treaty port and railway, and civil service exam quota per capita. Correlations between these controls and *New Youth* circulation are examined in the appendix. Standard errors in parentheses are clustered at the prefecture level; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3: Baseline Estimates: Impact of Post-19 $NewYouth$ Exposure on CCP Membership
Dependent Variable: CCP=1/0

New Youth exposure defined at	Prefecture level						County level					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
					Teams 1-4						Teams 1-4	
New Youth exposure, 1919-22	0.109**		0.128**	0.127**	0.127**	0.117**	0.153**		0.152**	0.181**	0.162**	0.172**
	(0.048)		(0.050)	(0.052)	(0.058)	(0.056)	(0.077)		(0.077)	(0.073)	(0.081)	(0.082)
New Youth exposure, 1915-18		-0.047	-0.081*	-0.055	-0.091	-0.094		-0.044	-0.038	-0.006	-0.103	-0.104
		(0.046)	(0.048)	(0.051)	(0.066)	(0.068)		(0.099)	(0.097)	(0.095)	(0.115)	(0.116)
Education: vocational				0.009	0.022	0.018				-0.066	-0.074	-0.078
				(0.046)	(0.047)	(0.047)				(0.067)	(0.075)	(0.076)
Education: college				-0.028	-0.003	-0.002				0.041	0.003	-0.005
				(0.062)	(0.073)	(0.075)				(0.138)	(0.159)	(0.179)
#CCP references				0.181***	0.171***	0.173***				0.179**	0.076	0.079
				(0.048)	(0.050)	(0.049)				(0.078)	(0.089)	(0.092)
Family: gentry						-0.015						-0.030
						(0.052)						(0.092)
Family: business						0.021						0.004
						(0.052)						(0.090)
Family economic status (1-3)						0.022						-0.011
						(0.039)						(0.059)
#Siblings						-0.001						-0.000
						(0.007)						(0.011)
Married						-0.072*						-0.022
						(0.043)						(0.078)
Having children						-0.012						-0.076
						(0.064)						(0.101)
Religious						-0.030						-0.069
						(0.051)						(0.090)
Smoking and drinking						0.042						0.042
						(0.087)						(0.143)
Prefecture FE	Y	Y	Y	Y	Y	Y						
County FE							Y	Y	Y	Y	Y	Y
Team FE				Y	Y	Y				Y	Y	Y
Birth year FE				Y	Y	Y				Y	Y	Y
Observations	689	689	689	689	530	530	689	689	689	689	530	530
R-squared	0.386	0.381	0.389	0.443	0.534	0.542	0.560	0.552	0.560	0.603	0.747	0.750

Note. Columns (1)–(4) and (7)–(10) include all cadets whereas Columns (5)–(6) and (11)–(12) focus on Team 1–4 on whom we have more information. Standard errors in parentheses are clustered at the home county level; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 4: Controlling for Prefecture FEs × Post 1919
Dependent Variable: CCP=1/0

New Youth exposure defined at	Prefecture level		County level	
	(1)	(2)	(3)	(4)
New Youth exposure, 1919-22	0.126**	0.125**	0.150*	0.181**
	(0.054)	(0.057)	(0.085)	(0.083)
New Youth exposure, 1915-18	-0.081	-0.053	-0.043	-0.006
	(0.050)	(0.053)	(0.101)	(0.100)
Prefecture FE	Y	Y		
County FE			Y	Y
Prefecture FE * 1919-22	Y	Y	Y	Y
Personal characteristics I		Y		Y
Team FE		Y		Y
Observations	689	689	689	689
R-squared	0.391	0.445	0.569	0.611

Note. Personal characteristics I include birth year fixed effects, vocational education, college, and number of CCP references. Standard errors in parentheses are clustered at the home county level; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 5: Always-New Youth vs. Not-always-New Youth
Dependent Variable: CCP=1/0

	(1)	(2)	(3)
Always-New Youth exposure, 1919-22	0.109**	0.126**	0.117**
	(0.049)	(0.050)	(0.056)
Not-always-New Youth exposure, 1919-22	0.121*	0.125*	0.151**
	(0.073)	(0.073)	(0.071)
New Youth exposure, 1915-1918		-0.078*	-0.049
		(0.047)	(0.050)
Team FE			Y
Birth year FE			Y
Personal characteristics I			Y
Observations	689	689	689
R-squared	0.388	0.390	0.445

Note. This table separates the exposures to those localities always with the *New Youth* and those with the *New Youth* entry and exit. Personal characteristics I include birth year fixed effects, vocational education, college, and number of CCP references. Standard errors in parentheses are clustered at the home county level; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 6: Impacts of Placebo Exposures: Assuming One Were in His Previous or Next Locations
Dependent Variable: CCP=1/0

	Relocation to the next location				Relocation to the previous location			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Teams 1-4				Teams 1-4
Placebo New Youth exposure, 1919-22	0.041		0.035	0.041	0.098		0.065	0.018
	(0.046)		(0.045)	(0.050)	(0.068)		(0.060)	(0.064)
Placebo New Youth exposure, 1915-18		0.033	0.024	0.010		0.128*	0.108*	0.062
		(0.039)	(0.038)	(0.049)		(0.067)	(0.061)	(0.063)
Prefecture FE	Y	Y	Y	Y	Y	Y	Y	Y
Team FE	Y	Y	Y	Y	Y	Y	Y	Y
Personal characteristics I	Y	Y	Y	Y	Y	Y	Y	Y
Personal characteristics II				Y				Y
Observations	689	689	689	530	689	689	689	530
R-squared	0.456	0.455	0.457	0.558	0.458	0.460	0.461	0.558

Note. Columns (1)–(4) assume that one were located in the previous prefecture according to his life event records whereas Columns (5)–(8) assume that one were located in the next prefecture.
Personal characteristics I include birth year fixed effects, vocational education, college, and number of CCP references; Personal characteristics II include family backgrounds (gentry, business), family economic status (1-3), number of siblings, marriage status, a dummy indicating whether they have children, religious belief, and smoking and drinking habit.
Standard errors in parentheses are clustered at the home county level; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 7: Other Pro-Communism Periodicals and Activist Groups
Dependent Variable: CCP=1/0

(a) New Youth exposure vs. Two other Communist periodicals

	(1)	(2)
New Youth, 1919-22 + no other two Communist mag.	0.116*	0.142**
	(0.061)	(0.063)
New Youth, 1919-22 + other two Communist mag.	0.134**	0.120**
	(0.055)	(0.057)
Prefecture FE	Y	Y
New Youth Exposure, 1915-18	Y	Y
Team FE		Y
Personal characteristics I		Y
Observations	689	689
R-squared	0.389	0.443

(b) New Youth exposure vs. Activist groups

	(1)	(2)
New Youth, 1919-22 + no Activist org.	0.103**	0.110**
	(0.049)	(0.051)
New Youth, 1919-22 + Activist org.	0.204***	0.178**
	(0.076)	(0.082)
New Youth Exposure, 1915-18	Y	Y
Prefecture FE	Y	Y
Team FE		Y
Personal characteristics I		Y
Observations	689	689
R-squared	0.392	0.444

Note. This table compares exposure to post-1919 *New Youth* and those also exposed to other Communist magazines and activist groups. Personal characteristics I include birth year fixed effects, vocational education, college, and number of CCP references. Standard errors in parentheses are clustered at the home county level; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 8: Impact on Future Political Outcomes

(a) SUR Model

	(1) 1924-1949		(2) 1927-1936		(3) 1924-1949	
	(1.1) Quit	(1.2) Sacrifice	(2.1) Quit	(2.2) Sacrifice	(3.1) Quit	(3.2) Sacrifice
(1) KMT with NY exposure	0.029	0.147***	-0.006	0.055	0.026	0.024
	(0.046)	(0.053)	(0.025)	(0.038)	(0.037)	(0.042)
(2) CCP without NY exposure	0.353***	0.304***	0.580***	0.249***	0.420***	0.214**
	(0.092)	(0.106)	(0.053)	(0.081)	(0.085)	(0.096)
(3) CCP with NY exposure	0.120**	0.550***	0.327***	0.413***	0.168***	0.424***
	(0.057)	(0.065)	(0.033)	(0.049)	(0.049)	(0.055)
(3)-(2)	-0.233***	0.246**	-0.253***	0.164*	-0.252***	0.210***
	(0.100)	(0.115)	(0.057)	(0.087)	(0.089)	(0.101)
NY Exposure, 1915-18	Y	Y	Y	Y	Y	Y
Prefecture FE	Y	Y	Y	Y		
Team FE	Y	Y	Y	Y		
Personal characteristics I	Y	Y	Y	Y	Y	Y
Observations	689	689	594	594	689	689
R-squared	0.356	0.376	0.588	0.417	0.063	0.125

(b) Multinomial Probit Model

	(1)CCP=0	(2) CCP=1 & Quit	(3) CCP=1 & Survival	(4) CCP=1 & sacrifice
New Youth exposure, 1919-1922	-0.111***	0.005	0.021	0.085***
	(0.037)	(0.020)	(0.016)	(0.031)
Pairwise comparisons:				
vs. (1) CCP=0		0.117**	0.132***	0.196***
		(0.049)	(0.045)	(0.063)
vs. (2) CCP=1 & Quit			0.015	0.080**
			(0.026)	(0.039)
vs. (3) CCP=1 & Survival				0.064*
				(0.036)
New Youth exposure, 1915-18	Y	Y	Y	Y
Personal characteristics I	Y	Y	Y	Y
Observations	689	689	689	689

Note. Table (a) compares the future outcomes of our studied individuals. The reference group is the KMT without post-1919 New Youth exposure. In Columns (1)-(2), personal characteristics I include birth year fixed effects, vocational education, college, and number of CCP references. Table (b) presents the marginal effects from multinomial probit estimations. Standard errors, indicated in parentheses, are reported with significance levels: *** for $p < 0.01$, ** for $p < 0.05$, and * for $p < 0.1$.

Online Appendix

Table of Contents

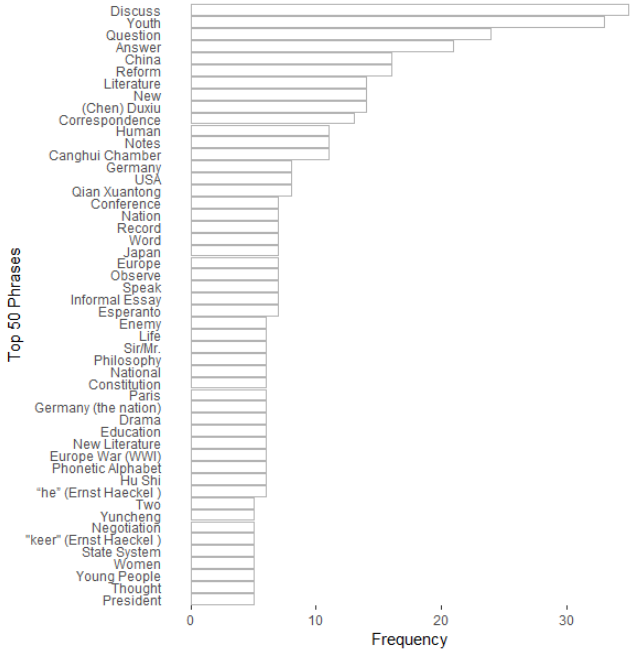
A Background and Data: More Results

A.1 Ideological Change of the *New Youth*: Top words

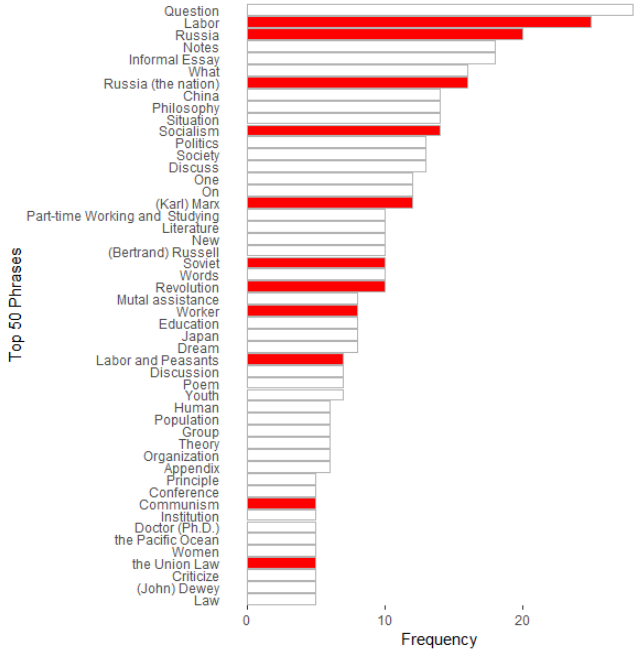
Figure A.1 presents the top words from the titles of *New Youth* articles. Panel (a) displays the pre-1919 articles, while panel (b) focuses on post-1919 articles. As shown in Panel (b), several words related to Communism, such as labor, Russia, and socialism, become prominent after 1919.

Figure A.1: Top words in the titles pre-post 1919

(a) Pre 1919



(b) Post 1919



A.2 More Personal Accounts on the Influence of the *New Youth*

In addition to Mao Zedong and Deng Xiaoping, many other Communist pioneers attribute the shaping of their ideas to the *New Youth*. Here, we provides a summary of their accounts, which reveal two patterns. First, the influence of the *New Youth* was widespread across various regions of China. Second, the *New Youth* served as a source of inspiration for other publications (which are also explored in this paper), and it was common for young people to read both *New Youth* and its satellite publications. Therefore, we further collect data on other pro-Communism periodicals in this era to help interpret our findings on post-1919 *New Youth*.

Peng Dehuai (in Hunan). According to Snow (1937), Peng Dehuai's reading experience is similar to Mao's. Snow notes that "Peng read Liang Qichao and Kang Youwei and many of the writers who had influenced Mao Zedong. For a time he had some interest in anarchism. In Chen Duxiu's *New Youth* he learned of socialism, and from that point he began to study Marxism."

Zhou En-lai (in Tianjin). According to Snow's account, Zhou En-lai informed him that he had read various Communist publications through the efforts of the *New Youth*. Snow quotes Zhou En-lai as stating, "I read translations of the *Communist Manifesto*; Kautsky's *Class Struggle*; and *The October Revolution*. These books were published under the auspices of the *New Youth*, edited by Chen Duxiu" (Snow 1937).

Zhang Guotao (in Beijing). Zhang Guotao mentioned in his memoir (Zhang 1971) that he was influenced by the *New Youth*, and that it served as an inspiration for other satellite magazines. He noted, "The circulation of the *New Youth* increased significantly, and every issue sold out immediately at Peking University. In December 1918, Chen Duxiu launched another politically-oriented publication called the *Weekly Review*, and in January 1919, Fu Sinian and Luo Jialun, who were fellow students at Peking University, published *New Tide*, a monthly literature magazine that responded to the New Culture Movement. These two magazines were considered as satellites of the *New Youth* at the time."

Yun Daiying (in Hubei). The diaries of Yun Daiying in the years before and after 1919 reveal that how he transitioned from supporting the cultural movement to embracing Marxist revolution. In his diary on September 9, 1919, Yun wrote, "I enjoy reading the *New Youth* and *New Tide* because they spread the message of freedom, equality, love, mutual aid, and labor"(Yun 1981).

Shi Cuntong (in Zhejiang). Yeh (1996) notes that the *New Youth* had a significant impact on Shi Cuntong, and the magazine was extremely popular among students in Hangzhou. She writes,

"In the fall of 1919, Shi Cuntong and Fu Binran, a fellow senior at First Normal, opened a student bookstore called the New Life Book Society, which aimed to disseminate New Culture publications in the province. The bestsellers on their stands were the *New Youth* magazine and the *Weekly Review*. The new tide of iconoclasm was sweeping over Hangzhou's students, with over four hundred copies of these two journals being sold weekly on the First Normal campus alone, which was approximately the school's total enrollment."

A.3 Summary Statistics and Data Sources

Table A.3: Data Sources and Summary Statistics

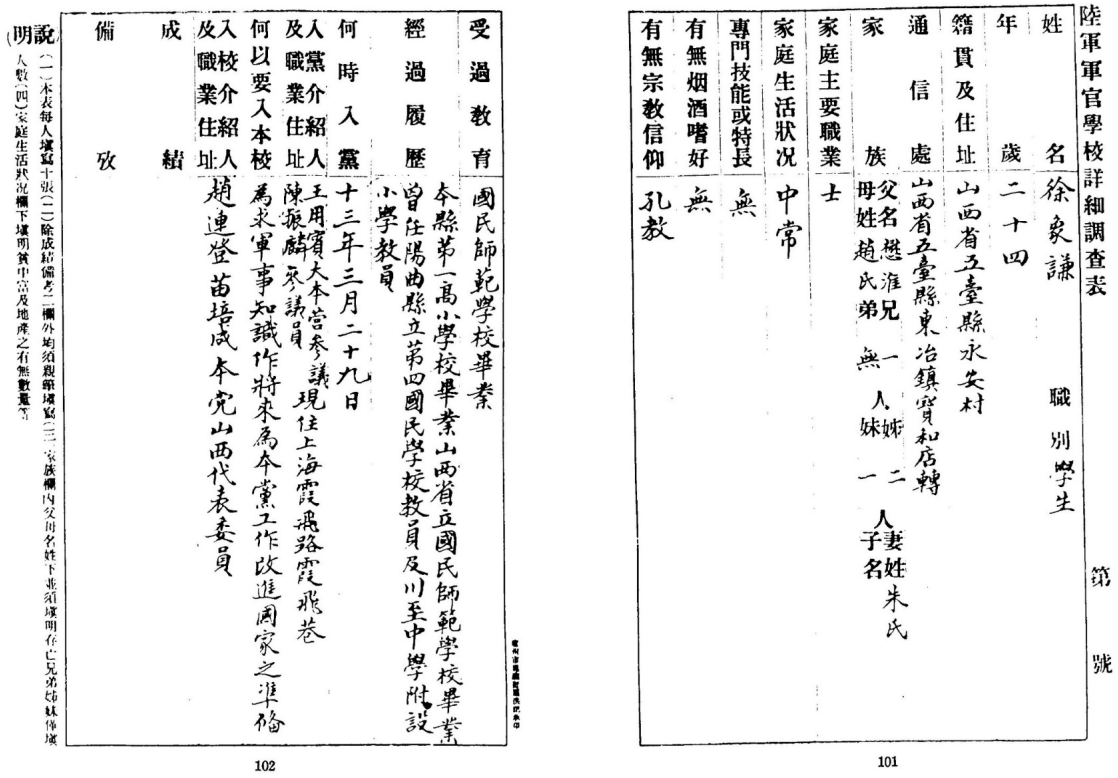
Variables	Data source	Observations	Mean	SD	Min	Max
CCP membership	A	689	0.171	0.377	0	1
Quit	A	689	0.160	0.367	0	1
Sacrifice	A	689	0.241	0.428	0	1
NY exposure, post-19	A, B	689	0.737	0.440	0	1
NY exposure, pre-19	A, B	689	0.552	0.498	0	1
NY exposure, post-19 (County)	A, B	689	0.530	0.499	0	1
NY exposure, pre-19 (County)	A, B	689	0.293	0.456	0	1
Just miss	A, B	689	0.125	0.331	0	1
Other Communist magazines	C	689	0.466	0.499	0	1
Activist Groups	D	689	0.376	0.485	0	1
Team	A	689	3.274	1.736	1	6
Birth year	A	689	1900.7	2.888	1886	1909
Education: middle school and above	A	689	0.956	0.204	0	1
Education: vocational	A	689	0.605	0.489	0	1
Education: college	A	689	0.107	0.310	0	1
#CCP references	A	689	0.180	0.549	0	5
Teams 1-4 only						
Family: gentry	A, E	530	0.249	0.433	0	1
Family: business	A, E	530	0.308	0.462	0	1
Family economic status (1-3)	A, E	530	1.634	0.538	1	3
# Siblings	A, E	530	3.383	2.430	0	16
Married	A, E	530	0.543	0.499	0	1
Having children	A, E	530	0.138	0.345	0	1
Religious	A, E	530	0.109	0.312	0	1
Smoking and drinking	A, E	530	0.051	0.220	0	1

Sources.: A. Chen (2017a), *Complete Records on the First-Cohort Cadets of the Whampoa Military School.*
B. The *New Youth.* Reprinted in 1970. Tokyo: Kyukoshoin.
C. Xue (ed. 2014), *Red Collections: Progressive Periodicals from 1915–1949.*
D. Zhang et al. (ed. 1979), *Voluntary Societies during the May Fourth Era.*
E. Central Executive Committee of the Kuomintang (1924), *Detailed Survey for the First to Fourth Teams of the First Cohort of the Military Academy.*

A.4 The Whampoa Cadets: A Biographical Dataset

We use two related sources of data to construct a biographical dataset on the first cohort of Whampoa cadets. First, Chen (2017a) compiled the biographies for all 689 cadets. Second, the Whampoa conducted a detailed survey for the 530 cadets that were directly recruited (Teams 1-4). Figure A.4(I) presents an example of the admission questionnaires.

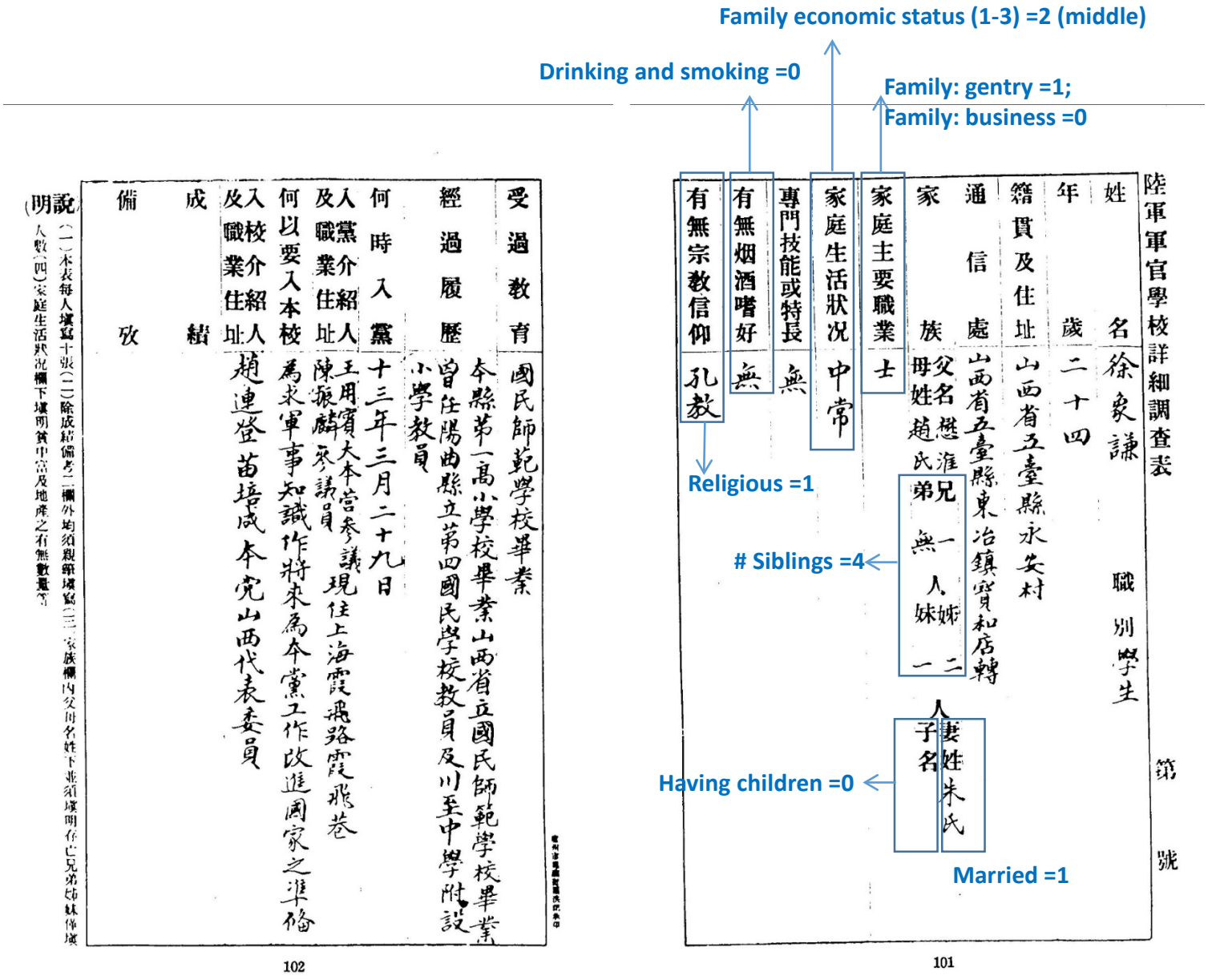
Figure A.4: I. The Admission Questionnaire of Xu Xiangqian



Figures A.4(II) and (III) provide an example that illustrates our coding process, explained below.

Figure A.4: II. An Example of Coding Based on the Biography

Figure A.4: III. An Example of Coding Based on the Detailed Survey



i. CCP Membership and Other Political Outcomes. Based on the cadets' biographies, we define three binary variables to indicate their CCP membership, whether they quit the party, and whether they died during the war. For example, Xu Xiangqian's biography shows that he joined the CCP in 1927 and became one of the ten marshals of the People's Liberation Army in 1955. Therefore, for Xu Xiangqian, CCP membership=1, quitting=0, and sacrifice=0. We highlight the corresponding coding in red in Figure A.4(II).

ii. Personal Characteristics I. We extract four sets of variables from the cadets' biographies to measure their basic personal characteristics, as illustrated in Figure A.4(II). First, we code the cadet's birth year. Second, based on their educational background, we construct three dummy variables to indicate whether the cadet received middle school education or above, vocational education, and college education. Third, we list references from the cadet's application to the Whampoa, check the party status of the references, and then calculate how many CCP references the cadet had. Fourth, we code the team to which the cadet was assigned. The variables related to Xu Xiangqian are highlighted in blue in Figure A.4(II).

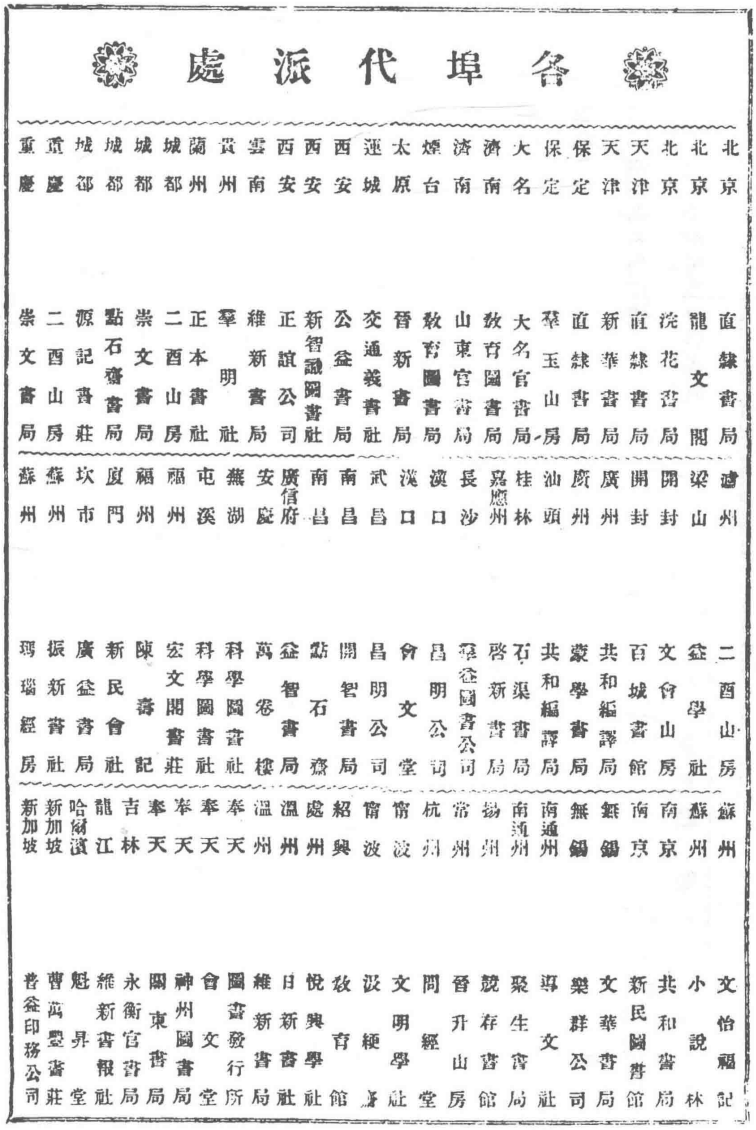
iii. Personal Characteristics II. We use the survey conducted by Whampoa to construct three additional sets of variables to further measure the personal characteristics of the cadets. First, we code the family background of the cadets. Specifically, we use two dummy variables to measure whether the cadet's family had an educational tradition and whether they ran a business. Another categorical variable is constructed to measure the cadet's family economic status, with categories 1 (poor), 2 (middle), and 3 (rich). Second, we collect more demographic information, including the number of siblings, whether the cadet was married, and whether the cadet had children. Third, we use two dummy variables to measure the cadet's religion and behavior traits, specifically, whether the cadet believed in any religion and whether the cadet had habits of drinking and smoking. The related coding for Xu Xiangqian is highlighted in blue in Figure A.4(III).

A.5 Coding Life History and Exposure to the *New Youth*

To determine whether an individual cadet was exposed to the *New Youth* and Communist ideology, we take three steps, as explained below.

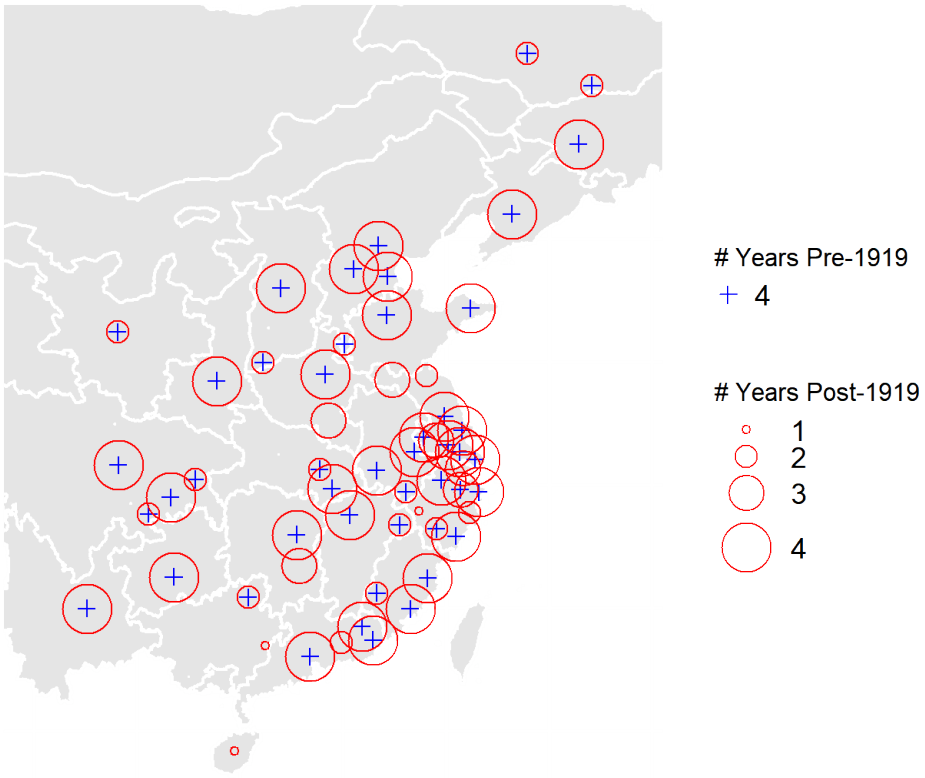
Step 1. Coding the Circulation of the *New Youth* Over Time. We first code the circulation locations of the *New Youth* magazine over time. Figure A.5(I) provides an example.

Figure A.5: I. An Example of the Circulation of the *New Youth*



We code the circulation locations for all volumes during 1915–22. Figure A.5(II) presents the geographic circulation of the *New Youth* over time. As shown, the *New Youth* was consistently available in certain prefectures, while there were entry and exit points in other prefectures. Not surprisingly, the former group tends to include larger cities.

Figure A.5: II. Circulation Locations of the *New Youth*



Step 2. Coding the Locations of the Cadets Over Time. To code the locations of the cadets over time, we use the information on their recorded life events from their biographies, focusing on time and location, as illustrated in Figure A.5(III). This allows us to construct an individual-year-level panel dataset with the cadets' location. Figure A.5(IV) presents the geographical distribution of all cadets in each year during 1915–22, showing that they were widely dispersed across China.

Figure A.5: III. Coding Life History

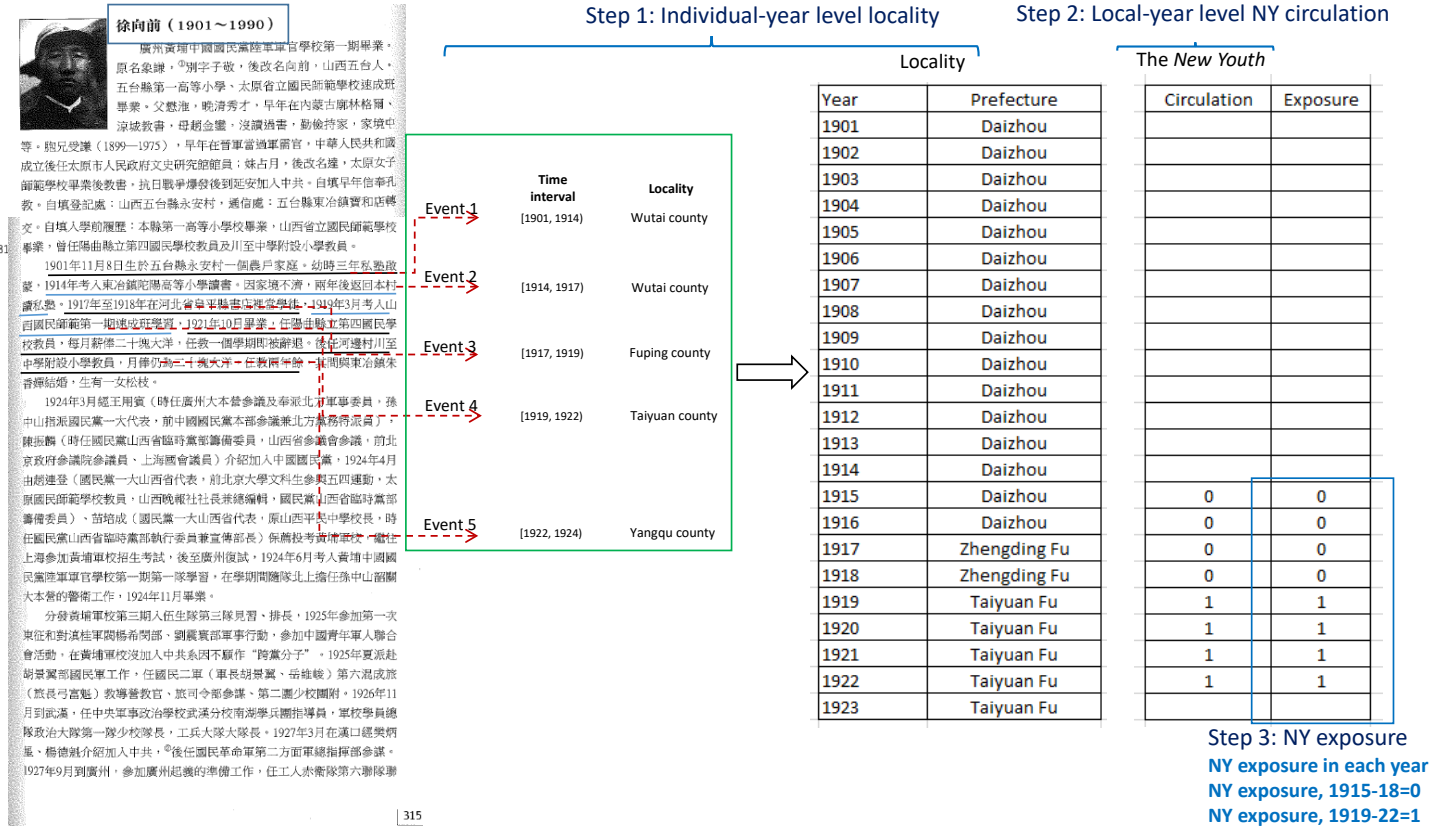
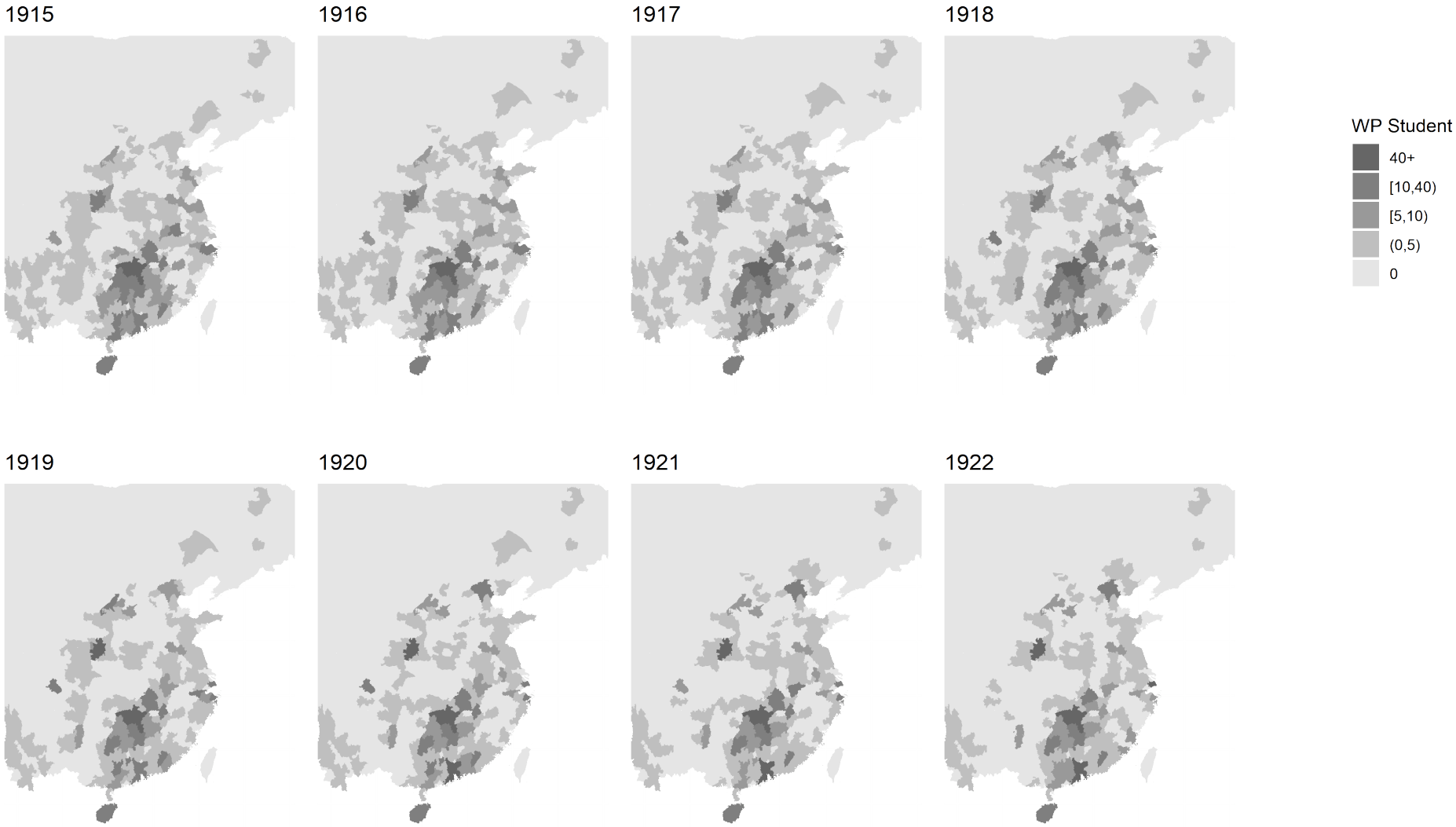
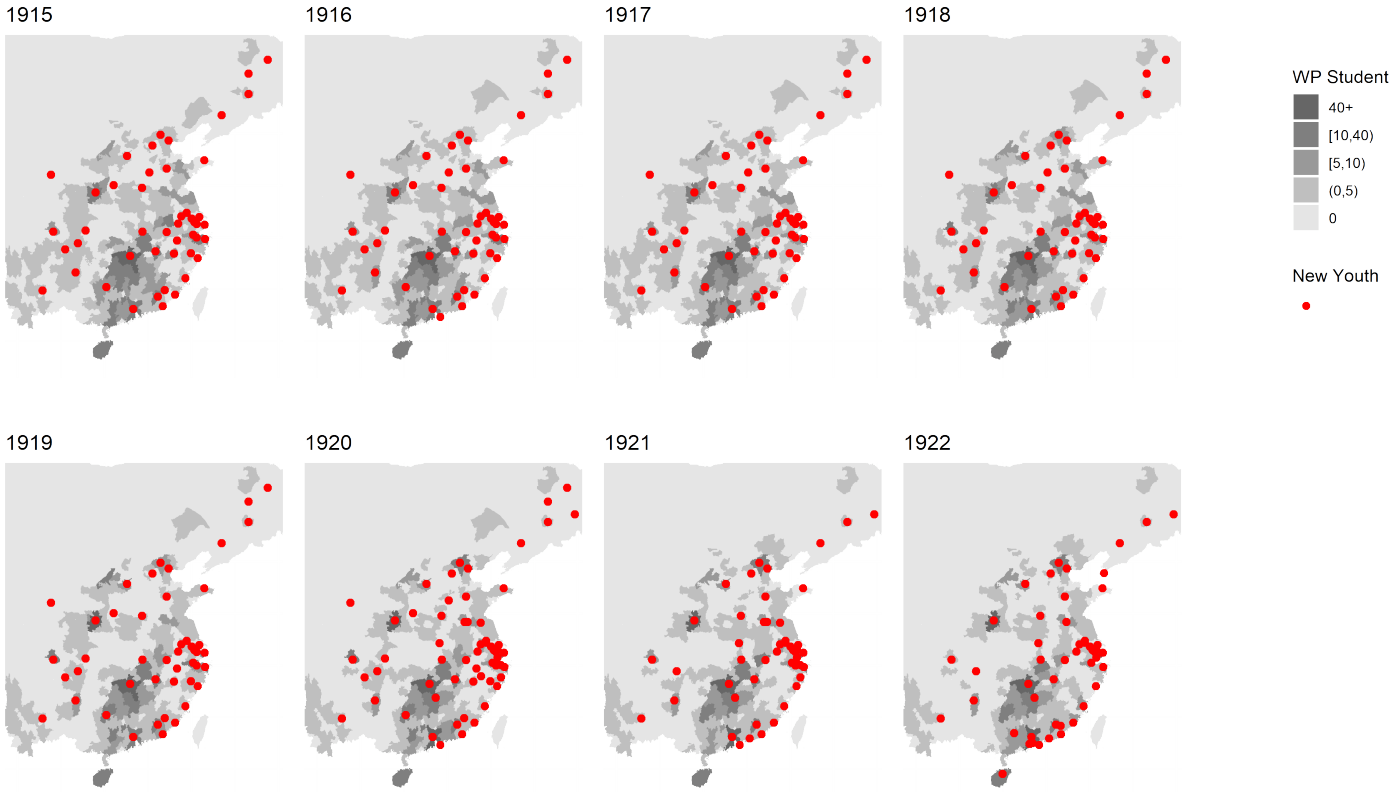


Figure A.5: IV. Cadet Location Year by Year



Step 3. Defining Exposure to the *New Youth*. To determine whether a cadet was exposed to the *New Youth* and Communist ideology, we combine information on the cadet's location and the *New Youth*'s circulation as shown in Figure A.5(V). If a cadet lived in a locality with *New Youth* circulation during a particular year (represented by the gray background and red point respectively), then we consider the cadet to be exposed to the *New Youth*. For example, using Xu Xiangqian as an example, he lived in Taiyuan during 1919–22, and since the *New Youth* was available in this location during this time, Xu was exposed to the *New Youth*.

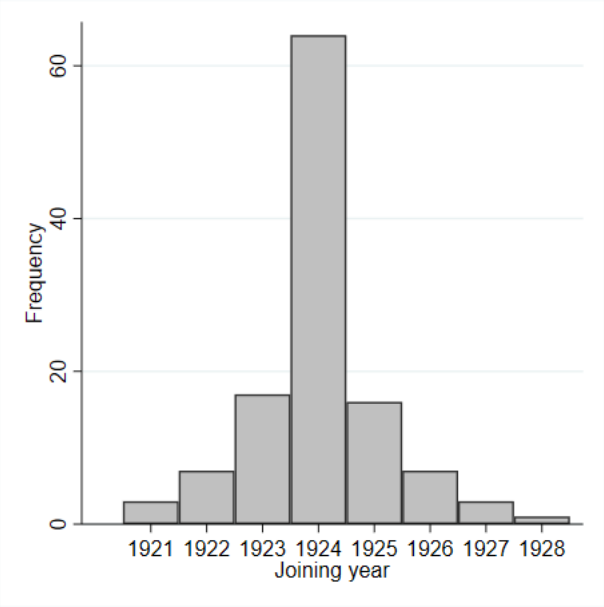
Figure A.5: V. Exposure to the *New Youth*



A.6 Years of Joining the CCP

Figure A.6 illustrates the timeline of CCP membership among the participants in our study. The graph indicates that all participants joined the party during its early years, with 77% joining during or after the Whampoa period. Because of limited access to party organizations in the early 1920s, few participants joined the party before Whampoa.

Figure A.6: The Distribution of CCP Years



A.7 Determinants of *New Youth* Circulation

We construct a prefecture-level panel dataset spanning two periods, one before and the other after 1919, to investigate the factors influencing the circulation of *New Youth*. Our dependent variable is the duration (in years) of *New Youth* circulation in each prefecture for both periods. We consider three sets of prefecture characteristics: (1) population and urbanization, (2) treaty port and railway presence that facilitated information distribution, and (3) political capital proxied by civil service exam quota per capita. As shown below, *New Youth* predominantly circulated in prefectures characterized by larger urban population, larger population, the presence of treaty port and railway, as seen in Columns (1) and (2). This pattern persisted consistently across both timeframes, showing no significant fluctuations over time, as indicated in Column (3).

Table A.7: Determinants of New Youth Circulation

Dependent variable:	Number of years with the New Youth circulation		
	(1)	(2)	(3)
Sample	1915-18	1919-22	(2)-(1)
City population	0.173*	0.162*	-0.011
	(0.088)	(0.084)	(0.012)
Population, 1910s	0.160**	0.202***	0.041
	(0.075)	(0.076)	(0.033)
Treaty port (dummy)	0.901***	0.832**	-0.069
	(0.342)	(0.329)	(0.108)
Railway (dummy)	0.366	0.564**	0.198
	(0.250)	(0.248)	(0.127)
Quota per million population	0.001	0.001	-0.000
	(0.000)	(0.000)	(0.000)
Observations	256	256	512
R-squared	0.403	0.502	0.063
Number of prefectures			256

Note. Standard errors in parentheses are clustered at the prefecture level; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B Additional Analyses

B.1 Heterogeneous Effects across Personal Characteristics

Table B.1 presents the interaction effects between post-1919 *New Youth* exposure and various personal characteristics. While there seems to be some complementarity between social networks and ideology exposure, there are no strong heterogeneous patterns in general.

Table B.1: Impact of Post-1919 *New Youth*: Heterogeneous Effects. Dependent Variable: CCP=0/1

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
NY exposure, 1919-22	0.122**	0.112*	0.118**	0.126**	0.117**	0.125**	0.117**	0.117**	0.116**	0.117**	0.118**	0.118**
	(0.055)	(0.058)	(0.055)	(0.055)	(0.056)	(0.056)	(0.058)	(0.056)	(0.056)	(0.056)	(0.055)	(0.056)
NY*1(Birth year>1900)	-0.116											
	(0.095)											
NY*Education: vocational		-0.075										
		(0.110)										
NY*Education: college			0.019									
			(0.132)									
NY*# CCP references				0.135*								
				(0.076)								
NY*Family: gentry					0.049							
					(0.087)							
NY*Family: business						0.151*						
						(0.091)						
NY*Family economic status							0.003					
							(0.085)					
NY*# siblings								0.000				
								(0.014)				
NY*Married									0.010			
									(0.090)			
NY*Having child										-0.022		
										(0.129)		
NY*Religious											0.044	
											(0.116)	
NY*Smoking and drinking												0.071
												(0.146)
NY exposure, 1915-18	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Prefecture FE, Team FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Personal charar. I, II	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	530	530	530	530	530	530	530	530	530	530	530	530
R-squared	0.545	0.543	0.542	0.546	0.543	0.546	0.542	0.542	0.542	0.542	0.542	0.542

Note. Standard errors in parentheses are clustered at the home county level; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B.2 Alternative Standard Errors

Robust standard errors are reported in parentheses, whereas two-way clustering standard errors at home county level and birth year level are presented in brackets. These standard errors are close to those clustered at the home county level.

Table B.2: Impact of Post-1919 *New Youth*: Various Standard Errors
Dependent Variable: CCP=0/1

New Youth exposure defined at	Prefecture level						County level					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
					Teams 1-4						Teams 1-4	
New Youth exposure, 1919-22	0.109		0.128	0.127	0.127	0.117	0.153		0.152	0.181	0.162	0.172
	(0.042)***		(0.044)***	(0.047)***	(0.054)**	(0.053)**	(0.064)**		(0.064)**	(0.070)***	(0.081)**	(0.083)**
	[0.046]**		[0.048]**	[0.050]**	[0.053]**	[0.051]**	[0.057]**		[0.057]**	[0.053]***	[0.048]***	[0.046]***
New Youth exposure, 1915-18		-0.047	-0.081	-0.055	-0.091	-0.094		-0.044	-0.038	-0.006	-0.103	-0.104
		(0.046)	(0.048)*	(0.051)	(0.059)	(0.061)		(0.080)	(0.079)	(0.085)	(0.101)	(0.103)
		[0.042]	[0.040]*	[0.044]	[0.052]	[0.050]*		[0.106]	[0.107]	[0.102]	[0.110]	[0.109]
Prefecture FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
County FE							Y	Y	Y	Y	Y	Y
Team FE				Y	Y	Y				Y	Y	Y
Personal characteristics I				Y	Y	Y				Y	Y	Y
Personal characteristics II						Y						Y
Observations	689	689	689	689	530	530	689	689	689	689	530	530
R-squared	0.386	0.381	0.389	0.443	0.534	0.542	0.560	0.552	0.560	0.603	0.747	0.750

Note. Personal characteristics I include birth year fixed effects, vocational education, college, and number of CCP references; Personal characteristics II include family backgrounds (gentry, business), family economic status (1-3), number of siblings, marriage status, a dummy indicating whether they have children, religious belief, and smoking and drinking habit. Robust standard errors are presented in parentheses and two-way clustering standard errors at the hometown level and the birth year level in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B.3 A Continuous Measure of *New Youth* Exposure

Our finding is robust to using a continuous measure of *New Youth* exposure, where we calculate the exposure as the share of years being exposed. The results are presented in Table B.3.

Table B.3: A Continuous Measure of New Youth Exposure
Dependent Variable: CCP=0/1

	(1)	(2)
		Teams 1-4
Average New Youth exposure, 1919-22	0.159***	0.161**
	(0.058)	(0.070)
Average New Youth exposure, 1915-18	-0.002	-0.068
	(0.069)	(0.081)
Prefecture FE	Y	Y
Team FE	Y	Y
Personal characteristics I	Y	Y
Personal characteristics II		Y
Observations	689	530
R-squared	0.464	0.564

Note. Personal characteristics I include birth year fixed effects, vocational education, college, and number of CCP references; Personal characteristics II include family backgrounds (gentry, business), family economic status (1-3), number of siblings, marriage status, a dummy indicating whether they have children, religious belief, and smoking and drinking habit. Standard errors in parentheses are clustered at the home county level; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B.4 Comparison Groups by Cohort

Our comparison group comprises of individuals who have resided in the prefectures ever with post-1919 *New Youth* exposure but missed it. We can divide the comparison group into (i) individuals who lived in these prefectures before 1919, and (ii) individuals who resided in the prefectures during the post-1919 period but narrowly missed the *New Youth* exposure. Within group (ii), we further separate those who arrived in a locality during 1919–22 and those who arrived in 1923, the latter of which missed the *New Youth* exposure due to national-level exit. As shown in Table B.4, when comparing with these different cohorts, our estimates are similar to our baseline one, suggesting that sorting of individuals is unlikely to be the driver of our main finding.

Table B.4: Comparing Groups by Cohort
Dependent Variable: CCP=1/0

	(1)	(2)	(3)	(4)
New Youth exposure, 1919-22	0.127**	0.127**	0.136**	0.137**
	(0.052)	(0.052)	(0.068)	(0.067)
(1) Been in the pref. with Post-19 NY before 1919		-0.018		
		(0.089)		
(2) Been in the pref. with Post-19 NY after 1919, but just missed			0.016	
			(0.071)	
(2.1) Been in the pref. with Post-19 NY during 1919-22, but just missed				0.014
				(0.095)
(2.2) Been in the pref. with Post-19 NY in 1923, and missed NY				0.017
				(0.070)
New Youth exposure, 1919-22 vs. group (1)		0.145		
		(0.103)		
New Youth exposure, 1919-22 vs. group (2)			0.120**	
			(0.059)	
New Youth exposure, 1919-22 vs. group (2.1)				0.123
				(0.094)
New Youth exposure, 1919-22 vs. group (2.2)				0.120*
				(0.063)
New Youth Exposure, 1915-18	Y	Y	Y	Y
Prefecture FE	Y	Y	Y	Y
Team FE	Y	Y	Y	Y
Other personal characteristics I	Y	Y	Y	Y
Observations	689	689	689	689
R-squared	0.443	0.443	0.443	0.443

Note. This table compares those with exposure and those who had been to prefectures with post-1919 *New Youth* but missed it due to life events. Personal characteristics I include birth year fixed effects, vocational education, college, and number of CCP references. Standard errors in parentheses are clustered at the home county level; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B.5 Measurement Error

We conduct three sets of analyses to make sure our findings are not driven by measurement error at the locality or individual levels.

i. Dropping Metropolitan Cities. Our estimates are similar or a bit larger than our baseline estimate if we drop the metropolitan cities from our sample.

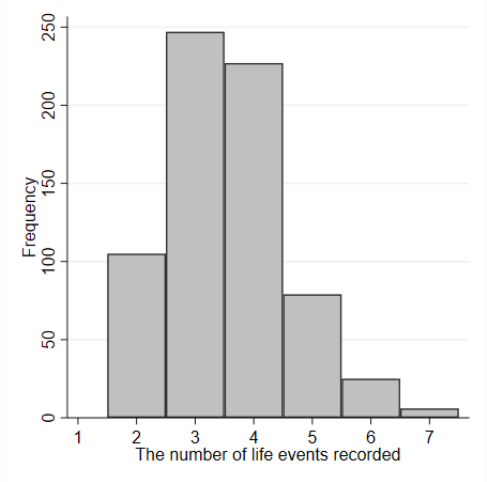
Table B.5: I. Dropping Big Cities
Dependent Variable: CCP=0/1

	All cadets				Teams 1-4			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Excluding who had been	Beijing	Shanghai	Guangzhou	All three	Beijing	Shanghai	Guangzhou	All three
CCP mean:	0.163	0.163	0.177	0.172	0.160	0.160	0.179	0.173
New Youth exposure, 1919-22	0.135***	0.133***	0.132**	0.212**	0.127**	0.126**	0.140**	0.199**
	(0.047)	(0.045)	(0.066)	(0.088)	(0.058)	(0.054)	(0.069)	(0.088)
New Youth exposure, 1915-18	Y	Y	Y	Y	Y	Y	Y	Y
Prefecture FE	Y	Y	Y	Y	Y	Y	Y	Y
Team FE	Y	Y	Y	Y	Y	Y	Y	Y
Personal char. I	Y	Y	Y	Y	Y	Y	Y	Y
Personal char. II					Y	Y	Y	Y
Observations	613	652	417	314	456	494	407	306
R-squared	0.438	0.445	0.592	0.563	0.547	0.558	0.602	0.577

Note. Personal characteristics I include birth year fixed effects, vocational education, college, and number of CCP references; Personal characteristics II include family backgrounds (gentry, business), family economic status (1-3), number of siblings, marriage status, a dummy indicating whether they have children, religious belief, and smoking and drinking habit. Standard errors in parentheses are clustered at the home county level; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

ii. Precision of Life Event Records. The major life events for our studied individuals—in their early 20s in 1924—are education and jobs. On average, each cadet has 3.6 recorded life event records. As Figure B.5 shows, the majority of the cadets (474 out of 689) have 3 or 4 recorded events. About 15% (105 out of 689) have 2 recorded events, and 16% (110 out of 689) have at least 5 recorded events.

Figure B.5: The Number of Life Events Recorded



To check whether our finding is affected by the number of events recorded, we examine whether our estimate varies by the number of records. This is not the case, as shown in Table B.5(II). Our

finding also holds if we remove those who have few or many events recorded.

Table B.5: II. Considering the Number of Recorded Life Events
Dependent Var.: CCP=0/1

	(1)	(2)	(3)	(4)	(5)	(6)
				Teams 1-4		
#Recorded events:		>2	<5		>2	<5
New Youth exposure, 1919-22	0.127**	0.165***	0.166***	0.112**	0.135**	0.153**
	(0.053)	(0.060)	(0.058)	(0.055)	(0.061)	(0.063)
#Recorded events * New Youth, 1919-22	0.011			-0.002		
	(0.045)			(0.049)		
#Recorded events:	0.052			0.065		
	(0.036)			(0.045)		
New Youth exposure, 1915-18	Y	Y	Y	Y	Y	Y
Prefecture FE	Y	Y	Y	Y	Y	Y
Team FE	Y	Y	Y	Y	Y	Y
Personal characteristics I	Y	Y	Y	Y	Y	Y
Personal characteristics II				Y	Y	Y
Observations	689	584	579	530	488	426
R-squared	0.455	0.489	0.448	0.554	0.574	0.599

Note. Personal characteristics I include birth year fixed effects, vocational education, college, and number of CCP references; Personal characteristics II include family backgrounds (gentry, business), family economic status (1-3), number of siblings, marriage status, a dummy indicating whether they have children, religious belief, and smoking and drinking habit. Standard errors in parentheses are clustered at the home county level; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

iii. Missing Years of Recorded Events. Some recorded events in the data did not include exact years, so we had to use assumptions to infer the missing years. Specifically, we used the beginning year of the next event or the school system to estimate the missing years. However, these assumptions may introduce some measurement error.

To check the validity of our method, we applied the same inference method to the entire sample by assuming that we only knew the birth year, end year, and life events for all individuals. We then compared the estimates from our original sample to the estimates from the sample with inferred years.

Table B.5 presents the results. Columns (1) and (3) show the original estimates, and Columns (2) and (4) show the estimates with inferred years. We found that the estimates were similar, which confirms that our findings are not driven by our inference method.

Table B.5: III. Validating the Inference Method
Dependent Var.: CCP=0/1

	(1)	(2)	(3)	(4)
			Teams 1-4	
New Youth exposure, 1919-22	0.127**		0.117**	
	(0.052)		(0.056)	
New Youth exposure, 1919-22		0.110**		0.123**
(Based on events only)		(0.056)		(0.058)
NY exposure, 1915-18	Y	Y	Y	Y
Prefecture FE	Y	Y	Y	Y
Team FE	Y	Y	Y	Y
Personal characteristics I	Y	Y	Y	Y
Personal characteristics II			Y	Y
Observations	689	689	530	530
R-squared	0.443	0.440	0.542	0.542

Note. Personal characteristics I include birth year fixed effects, vocational education, college, and number of CCP references; Personal characteristics II include family backgrounds (gentry, business), family economic status (1-3), number of siblings, marriage status, a dummy indicating whether they have children, religious belief, and smoking and drinking habit. Standard errors in parentheses are clustered at the home county level; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B.6 Heterogeneous Circulation Locations

To gauge the reading rate, we examine heterogeneity across circulation locations in three ways. First, we separate the prefecture-level post-1919 *New Youth* into a breakdown of within a county and outside the county. Columns (1)–(2) of Table B.6 show little impact if the post-1919 *New Youth* was available outside of a cadet's located county.

Second, we separate the circulation county into small and big cities, defined by a threshold of urban population size of 100,000. Using the same specification as our baseline, we obtain estimates of 0.26 for small cities and 0.13 for big cities (Columns (3)–(4)).

Finally, a subsample of the circulation points were bookstores located in schools. Based on the life history of the cadets, we know whether a cadet was studying in those schools. When separating this very local exposure from others, we find that the coefficient of this very local exposure is 0.26 whereas that of the rest is 0.15 (Columns (5)–(6)).

Table B.6: The Impacts of Ideology Exposure across Localities
Dependent Var.: CCP=0/1

	(1)	(2)	(3)	(4)	(5)	(6)
New Youth (NY) exposure, 1919-22 (pref. level)						
different county with the NY	0.040	0.051	0.038	0.049	0.032	0.044
	(0.050)	(0.054)	(0.049)	(0.054)	(0.050)	(0.054)
same county with the NY	0.160***	0.161***				
	(0.056)	(0.057)				
same county with the NY: small city			0.251***	0.260***		
			(0.088)	(0.080)		
same county with the NY: big city			0.135**	0.131**		
			(0.059)	(0.061)		
same county with the NY: same school					0.250*	0.263*
					(0.149)	(0.155)
same county with the NY: others					0.151***	0.152***
					(0.058)	(0.058)
NY Exposure, 1915-18	Y	Y	Y	Y	Y	Y
Prefecture FE	Y	Y	Y	Y	Y	Y
Team FE		Y		Y		Y
Birth year FE		Y		Y		Y
Personal characteristics I		Y		Y		Y
Observations	689	689	689	689	689	689
R-squared	0.400	0.451	0.403	0.455	0.400	0.452

Note. Personal characteristics I include birth year fixed effects, vocational education, college, and number of CCP references. Standard errors in parentheses are clustered at the home county level; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B.7 Possible Spillover Effects

We investigate potential spillover effects by examining two types of social connections among the cadets in our study. We observe some indications of spillover effects when cadets share the same reference, but we find that our primary result remains robust even when accounting for these possible effects.

Table B.7: Allowing for Spillover Effects
Dependent Variable: CCP=0/1

	(1)	(2)	(3)	(4)
New Youth exposure, 1919-22	0.125**	0.133**	0.125**	0.131**
	(0.052)	(0.054)	(0.052)	(0.052)
Post-1919 NY exposure of hometown (county) fellows	0.002			
	(0.002)			
Post-1919 NY exposure of hometown (prefecture) fellows		0.002		
		(0.001)		
Post-1919 NY exposure of hometown (province) fellows			-0.001	
			(0.001)	
Post-1919 NY exposure of connected cadets (via references)				0.005*
				(0.003)
New Youth exposure, 1915-18	Y	Y	Y	Y
Prefecture FE	Y	Y	Y	Y
Team FE	Y	Y	Y	Y
Personal characteristics I	Y	Y	Y	Y
Observations	689	689	689	689
R-squared	0.445	0.445	0.444	0.446

Note. Personal characteristics I include birth year fixed effects, vocational education, college, and number of CCP references. Standard errors in parentheses are clustered at the home county level; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B.8 Results from Cohorts 2–4

Table B.8 indicates that a comparable pattern exists for 700 chosen cadets from the second through fourth cohorts of the Whamphoa. Since there are fewer personal characteristic variables available for these cadets than for those in the first cohort, we have included only the essential variables in the analysis, allowing us to compare the results with those for the first cohort.

Table B.8: Results for Other Cohorts
Dependent Variable: CCP=0/1

Exposure defined at	Cohort 1				Cohorts 2-4			
	Prefecture level		County level		Prefecture level		County level	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
New Youth exposure, 1919-22	0.128**	0.118**	0.152**	0.158**	0.129***	0.128**	0.173**	0.240***
	(0.050)	(0.052)	(0.077)	(0.073)	(0.050)	(0.057)	(0.080)	(0.088)
New Youth exposure, 1915-18	-0.081*	-0.061	-0.038	-0.015	-0.005	-0.031	-0.089	-0.170
	(0.048)	(0.049)	(0.097)	(0.093)	(0.063)	(0.069)	(0.105)	(0.111)
Education: vocational		0.014		-0.050		0.001		-0.115
		(0.041)		(0.062)		(0.075)		(0.101)
Education: college		-0.029		0.035		-0.099		-0.232**
		(0.063)		(0.134)		(0.081)		(0.093)
#CCP references		0.178***		0.163**		0.522***		0.637***
		(0.046)		(0.069)		(0.097)		(0.108)
Prefecture FE	Y	Y	Y		Y	Y		
County FE			Y	Y			Y	Y
Birth year FE		Y		Y		Y		Y
Cohort FE								Y
Observations	689	689	689	689	700	700	700	700
R-squared	0.389	0.442	0.560	0.601	0.383	0.452	0.587	0.664

Note. Personal characteristics I include birth year fixed effects, vocational education, college, and number of CCP references. Standard errors in parentheses are clustered at the home county level; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B.9 Timing of Joining the CCP

Table B.9 investigates the year of CCP membership to aid in interpreting the impact of post-1919 *New Youth* exposure. The results indicate that the primary effect occurs during the period of 1924–1928, indicating that the influence of ideological exposure is not likely to be a result of local CCP recruitment.

Table B.9: Timing of Joining the CCP
Dependent Var.: CCP=0/1

	(1)	(2)	(3)	(4)
Time of Joining the CCP	1921–1923		1924–1928	
New Youth exposure, 1919-22	0.046*	0.030	0.092*	0.092*
	(0.024)	(0.022)	(0.049)	(0.050)
#CCP references	0.135***	0.019	0.066	0.065
	(0.040)	(0.020)	(0.048)	(0.044)
NY exposure, 1919-22 × #CCP references		0.146***		0.001
		(0.043)		(0.073)
New Youth exposure, 1915-18	Y	Y	Y	Y
Prefecture FE	Y	Y	Y	Y
Team FE	Y	Y	Y	Y
Other personal characteristics I	Y	Y	Y	Y
Observations	689	689	662	662
R-squared	0.510	0.524	0.412	0.412

Note. Personal characteristics I include birth year fixed effects, vocational education, college, and number of CCP references. Standard errors in parentheses are clustered at the home county level; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B.10 Potential Social Interactions

One alternative hypothesis is that post-1919 *New Youth* exposure primarily reflects social interactions independent of ideas. We examine this hypothesis as follows. For each cadet, we know who co-located with him before joining the Whampoa and calculate the average post-1919 *New Youth* exposure ratio among his co-locaters. If post-1919 *New Youth* captured primarily social interactions rather than ideas, we would expect a strong heterogeneous pattern: After arriving at Whampoa, those who were more likely to share experience with his exposed co-locaters would be more likely to join the CCP. As shown in Table B.10, this is not supported by the data.

Table B.10: Potential Social Interactions
Dependent Var.: CCP=0/1

	(1)	(2)	(3)
Ave. Post-19 NY exposure ratio within co-locators			-0.051
× New Youth exposure, 1919-22			(0.300)
New Youth exposure, 1919-22	0.119***	0.136***	0.136***
	(0.042)	(0.051)	(0.052)
Ave. Post-19 NY exposure ratio within co-locators	-0.105	-0.210	-0.191
	(0.085)	(0.157)	(0.179)
Prefecture FE		Y	Y
New Youth exposure, 1915-18	Y	Y	Y
Team FE	Y	Y	Y
Other personal characteristics I	Y	Y	Y
Observations	689	689	689
R-squared	0.123	0.445	0.445

Note. Personal characteristics I include birth year fixed effects, vocational education, college, and number of CCP references. Standard errors in parentheses are clustered at the home county level; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B.11 Assignments in Whampoa and the National Revolutionary Army

To check whether our finding on political outcomes is primarily driven by assignment, we examine whether post-1919 *New Youth* exposure is associated with political assignments. First, we examine whether individuals who were exposed to post-1919 *New Youth* were more or less likely to hold leadership positions (captain or vice-captain) in the Whampoa Academy. In Column (1) of Table B.11(I), we observe no significant correlation between the two variables. We also analyze the post-graduation assignments of Whampoa cadets in the National Revolutionary Army. The First Army was under the strict control of Chiang Kai-shek, the leader of the KMT, while the Fourth Army was more influenced by the CCP. Our findings indicate that post-1919 *New Youth* exposure is not correlated with these assignments either (Columns (2) and (3)).

When include these political assignments as control variables and find that our finding on sacrifice and quitting still holds, as shown in Table B.11(II).

Table B.11: I. Political Assignments as the Outcome

	(1)	(2)	(3)
	Leadership	Assignment in the National Revolutionary Army	
	at Whampoa	The 1st Army	The 4th Army
New Youth exposure, 1919-22	0.003	-0.073	-0.024
	(0.029)	(0.069)	(0.036)
New Youth exposure, 1915-18	Y	Y	Y
Prefecture FE, Team FE	Y	Y	Y
Personal characteristics I	Y	Y	Y
Observations	689	689	689
R-squared	0.526	0.347	0.306

Table B.11: II. Controlling for Political Assignments

	(1) 1924-1949		(2) 1927-1936	
	Quit	Sacrifice	Quit	Sacrifice
(1) KMT with New Youth exposure, 1919-22	0.045	0.153***	-0.002	0.054
	(0.044)	(0.053)	(0.025)	(0.038)
(2) CCP without New Youth exposure, 1919-22	0.400***	0.317***	0.586***	0.253***
	(0.090)	(0.106)	(0.053)	(0.081)
(3) CCP with New Youth exposure, 1919-22	0.146***	0.557***	0.335***	0.418***
	(0.055)	(0.065)	(0.033)	(0.050)
Leadership at Whampoa	-0.090	0.172**	-0.018	0.066
	(0.068)	(0.081)	(0.039)	(0.059)
The First Army	0.185***	0.051	0.045***	-0.007
	(0.028)	(0.033)	(0.016)	(0.025)
The Fourth Army	-0.005	0.066	-0.002	-0.028
	(0.055)	(0.065)	(0.032)	(0.048)
(3)–(2)	-0.253***	0.239**	-0.251***	0.165*
	(0.097)	(0.115)	(0.057)	(0.087)
New Youth exposure, 1915-18	Y	Y	Y	Y
Prefecture FE	Y	Y	Y	Y
Team FE	Y	Y	Y	Y
Personal characteristics I	Y	Y	Y	Y
Observations	689	689	594	594
R-squared	0.396	0.382	0.593	0.418

Note. Personal characteristics I include birth year fixed effects, vocational education, college, and number of CCP references. Standard errors in parentheses are clustered at the home county level; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.