Nation-building and mass migration: Evidence from Mandatory Palestine *

Laura Panza, Yanos Zylberberg
Discussion Paper 25/796

23 September 2025

School of Economics

University of Bristol
Priory Road Complex
Bristol
BS8 1TU
United Kingdom

University of BRISTOL
School of Economics

Nation-building and mass migration: Evidence from Mandatory Palestine*

Laura Panza Yanos Zylberberg

September 23, 2025

Abstract

This paper examines the grassroots of nation-building in times of mass migration. We study the emergence of cohesive communities and societal leadership within the scattered, diverse Jewish settlements of Mandatory Palestine between 1920 and 1947. Our empirical strategy relies on a new "frontier expansion" algorithm to predict the dynamics of Jewish settlement creation, which we combine with migrant characteristics in a shift-share design to isolate exogenous variation in the local composition of settlers across locations. We find that: (i) leaders who played a crucial role in shaping the early state of Israel emerged from *diverse* communities; (ii) these communities were more cohesive and maintained better relationships with Arab neighbors; and (iii) these effects are predominantly observed in kibbutzim, i.e., integrated settlements with communal lifestyle. Further evidence suggests that these diverse, tight-knit communities were facing and addressing nation-building challenges at a local level, e.g., setting up institutions to foster a shared identity.

*Panza: University of Melbourne, CEPR, email: laura.panza@unimelb.edu.au; Zylberberg: University of Bristol, CEPR, CESifo, email: yanos.zylberberg@bristol.ac.uk. We are grateful to the University of Melbourne and the University of Bristol for research support. We would like to thank Sascha Becker, Matteo Cervellati, Quoc-Anh Do, Pauline Grosjean, Tim Guinnane, Andrea Ichino, Paul Raschky, Valeria Rueda, Yannay Spitzer, Jordi Vidal-Robert, and Hans-Joachim Voth, as well as conference participants at The Economic History of Middle Eastern Countries 1500–2020 (2022), ASREC (Harvard, 2023), ASREC (Monash, 2023), Melbourne PolEcon/PolSci Workshop (ACU, 2023), AUSCLIO (Adelaide, 2023), ASE Annual Meeting (Napoli, 2023), and EHA Meeting (Sacramento, 2024), and seminar participants at Bologna, Bristol, EUI, Monash, Nottingham, Padova, and Utrecht for useful comments and suggestions. We are grateful to Susan Staszewski for outstanding research assistance and for retrieving exceptional archival sources. The research also benefited from the careful work of Ofer Pharhi and Tomer Mey. Panza acknowledges the support of the ARC DECRA grant DE50100809. Zylberberg acknowledges support of the ANR/ESRC/SSHRC, through the ORA grant ES/V013602/1 (MAPHIS: Mapping History). The usual disclaimer applies.

Wars, (de)colonization, and large population movements often lead to the formation of new societies that must grapple with weak institutions, inherited grievances, and differences in preferences, values, or languages. A well-studied case is the United States and the role of mass migration and settlement formation in shaping the American identity (Fouka et al., 2022; Bazzi et al., 2020). Another compelling example is Indonesia, whose islands with different cultures and languages achieved unity through autocratic regimes, the imposition of a common language, and resettlement programs (Bazzi et al., 2019). In both cases, as in many others, the process of nation-building feeds off more localized dynamics. A fascinating, yet understudied, context in which the formation of a nation ended up being tied to the making of local communities is Palestine under the British Mandate. There, Jewish migrants settled in frontier agricultural communities in a territory lacking formal institutions and shaped by tensions with the local Arab population (Shafir, 1996). This paper focuses on such context of sustained Jewish migration to understand the grassroots of nation-building in times of mass migration.

We study two key dimensions of nation-building within the scattered and diverse Jewish communities of Mandatory Palestine between 1920 and 1947: social cohesion and the development of societal leaders.[1] We develop a "frontier expansion" algorithm which (dynamically) identifies the rural locations within Mandatory Palestine that were most likely to host a new settlement and thus likely to host newly-arrived immigrants to Palestine. Coupled with the shifting composition of immigrant waves, this approach provides exogenous variation in settler diversity across the many small communities established during this period. We find that *diverse* settlements—those composed of settlers from varied birthplaces—were more cohesive, maintained better relations with nearby Arab villages, and were more likely to produce future societal leaders. These effects are predominantly observed in kibbutzim, where communal structures, local democracy, and shared educational systems facilitated cohesion and collective decision-making. These communities acted as social laboratories for local nation-building, developing institutions to manage diversity and foster unity—as evidenced by a systematic analysis of minutes covering hundreds of meetings organized in these settlements. In effect, while such pre-state institutional models were not adopted at the national level, they had a marked impact on nation-building through the future societal leaders who were nurtured within these "institutionally-pro-active" communities.

Our empirical analysis relies on measuring fine-grained heterogeneity across the Jewish communities of Palestine under the British Mandate: we create a novel dataset

[1]Due to data limitations (and better administrative coverage of the Jewish sector), we are not able to shed much light on the evolution of Arab settlements. We capture their interaction with Jewish settlements during that period from the analysis of meeting minutes taking place in Jewish communities and from pacts formed across communities during the First Israeli-Arab War (1947–1949).

of settlers' characteristics across 316 Jewish settlements, which we complement with data on their local institutional framework, first-nature geography (e.g., natural amenities), and second-nature geography (e.g., transportation infrastructure, market access). We gather an array of diverse data sources to best characterize community cohesion, relations with Arab villages, and the production of societal leaders who would be instrumental in the early state of Israel: biographies of politicians, military commanders, as well as cultural figures who would have prominent roles after 1948—coupled with 3,340 minutes of Knesset sessions (1949–1977); more than 6,000 pages of minutes covering local meetings between 1930–1947 (e.g., discussing social cohesion, absorption of immigration, relations with Arab neighbors); 3,400 photographs of these early communities; and historical reports and archives on the 1947–1949 war.

One empirical challenge is that the composition of Jewish settlements in Mandatory Palestine might have been shaped by central authorities, or by a few individuals or families with strong leadership capabilities, thus raising endogeneity concerns. To address this issue, we develop a new "frontier expansion" algorithm which predicts the dynamics of settlement creation. Intuitively, new settlements would appear in locations with favorable natural amenities and at a reasonable distance from existing settlements (i.e., not too close and not too far, typically along trading routes). Our "frontier expansion" algorithm flexibly captures this intuition and predicts the timing and location of settlement creation in an iterative fashion. We then use aggregate, time-varying statistics on the characteristics of new immigrants (as a "shift"), combine them with the previous procedure to allocate these newly-arrived migrants across space (as a "share"), and use the resulting prediction as exogenous variation in the local composition of settlers across settlements. This empirical strategy relies on a recent literature discussing identification in shift-share designs (Borusyak et al., 2022; Goldsmith-Pinkham et al., 2020). Such literature shows that identification would either ensue from numerous, dispersed and quasi-random shares or from numerous and quasi-random shifts: our empirical strategy attempts to isolate quasi-random variations in both objects, but the shifts (the composition of migrant inflows, driven by push factors at origin—see Buggle et al., 2023) are more plausibly exogenous to the heterogeneous local conditions.

We find that an important factor explaining (i) the rise of societal leaders, (ii) community cohesion, and (iii) relationships with Arab neighbors is local *diversity*, as captured by settlers' regions of origin.[2] Diverse communities were much more likely to produce the

[2]Our proxy for settlement diversity, as measured in a "census" of settlements (1945), is a Herfindahl index computed across 8 large regions of origin chosen by surveyors at that time (e.g., "Eastern Europe"), possibly reflecting survey costs (more precise origins were reported in earlier, more formal censuses) or the fact that Jewish identities did not coincide with country borders (e.g., within the Pale of Settlement). While the Jewish immigrants during the Mandate period were predominantly from Europe, they had different national, linguistic, and social backgrounds. Moreover, the Jewish population in Mandatory Palestine

future societal leaders which would be instrumental to the functioning of the early state of Israel. This "diversity" effect is predictive of the rise of leaders, irrespective of their political obedience or region of origin—with the exception of Mizrahi Jews from Yemen and Iraq.[3] The "diversity" effect however mostly predicts the emergence of constructive leaders, which are invested in community cohesion within the later state of Israel—a characterization based on their later rhetoric and thematic emphasis at the Knesset. Diversity also predicts within-settlement cohesion as inferred from (i) a visual recognition algorithm based on 3,400 photographs covering more than 40 settlements, and (ii) the systematic analysis of meeting minutes spanning 6,000 pages of discussions across 107 settlements. Finally, diverse communities appear to have been better integrated with their Arab neighbors, as suggested by the tone of meeting discussions, records of civilian casualties, and documented pacts of peaceful coexistence during the first Arab-Israeli War.[4] These patterns did not extend to centrally planned military operations that prescribed specific targets for attack and depopulation, indicating that diversity fostered positive interactions only in instances where communities retained local agency.

Our baseline specification controls for elements that enter our frontier expansion algorithm (first-nature geography, second-nature geography, polynomials in the year of settlement creation), as well as sub-district fixed effects, population deciles, and factors likely to generate tensions with neighboring Arab villages. In principle, even conditioning on the geography of settlement in Palestine, diverse settlements host different people and might thus differ along other important, mitigating characteristics, such as language, religiosity, occupations of settlers, economic development, settlement organization, and overall immigration incidence. Using balance tests and controlling for these important confounders in a comprehensive sensitivity analysis, we show that the diversity effect is not mitigated by higher economic development or human capital, different structures reflecting different ideological inclinations, or the prevalence of certain settlers' groups.

To better understand the mechanics of community building in diverse settlements, we conduct a heterogeneity analysis focusing on a distinctive feature of the "Jewish

also included a non-negligible Oriental minority, predominantly from Yemen and Iraq. Note that our baseline shift-share instrument relies instead on a Herfindahl index computed from *countries* of origin as documented in immigration cards filled by the successive migration waves—a modeling choice that we submit to an array of robustness checks.

[3]We run a number of robustness checks to probe the discretionary nature of our biographical selection and allocation of leaders across communities: we consider an alternative source listing 6,000 leaders, the *Encyclopedia of the Founders and Builders of Israel*; and we use alternative allocation procedures accounting for mobility, the age at which leaders resided in each settlement, whether they were born in Palestine or immigrated, and whether they were involved in Zionist organizations before migration.

[4]Such localized efforts to reduce violence may have had minimal impact on the broader trajectory of the war. Notably, however, these community-driven initiatives highlight the varied stances of Jewish communities concerning coexistence with the Arab population. We find evidence of such contrasting views in our minutes of meetings (e.g., "doves" in Ein Hayam and "hawks" in Ashdot Ya'akov).

frontier": the type of local institutions governing these communities. We find that the diversity effect observed across our main outcomes is mostly confined to kibbutzim—i.e., integrated, democratic settlements with communal lifestyle and strong collective institutions. We shed light on these social laboratories with archival work and qualitative evidence. A systematic analysis of settlement meeting transcripts using a large language model reveals that more diverse kibbutzim devoted greater attention to diaspora assimilation, while spending less time on issues of trust and cohesion, as well as on (negative) relations with neighboring Arab villages. A more narrative approach reveals that nation-building challenges were projected onto local communities. Like other organizational forms—moshavim, urban centers, and other settlements—diverse kibbutzim faced significant coordination problems, including the need to organize schooling, work, redistribution, and local public services, and to navigate complex relationships with neighboring Arab communities. However, these diverse kibbutzim managed to address these challenges more effectively by building their own local institutions from the ground up, while lacking a natural prior model (e.g., from settlers' places of origin) and contending with internal diversity in preferences and values. This endeavor—dissolving previous group ties and building a new common identity—would be replicated at a nationwide level to shape the early state of Israel, with many societal leaders having been formed and trained within these local laboratories.

The main contributions of this paper are to highlight a positive role for diversity in times of mass migration and to identify a novel role for grassroots institutions in fostering social cohesion and nurturing societal leadership. The research hinges on: the fascinating context of an informal state arising from scattered settlements; emblematic communal organizations whose direct impact on the development of Israel has been documented, but whose indirect effect through the nurturing of leaders is less known; unconventional data sources; and the resulting, novel measures of community building.

Our work closely relates to the literature exploring the origins of shared identity in fragmented countries, particularly regarding societal cohesion (see Rohner and Zhuravskaya, 2024, for a comprehensive review). Recent studies include Bazzi et al. (2019) on inter-ethnic integration in Indonesia, Fouka et al. (2022) on assimilation in the age of mass migration in the United States, or Carlitz et al. (2024) on the role of schooling in nurturing cohesion and political identities (echoing Alan et al., 2021; Billings et al., 2021).[5] Our approach differs from most recent work in both measurement and context. Standard indicators of group cohesiveness in diverse societies—such as spoken language,

[5]The idea of a unified American identity is further discussed and challenged in: Bazzi et al. (2020), who highlight the emergence of a frontier culture of individualism—an effect not mirrored in the "Jewish frontier" in Mandatory Palestine; and Giuliano and Tabellini (2020) on the ideological transmission by immigrants to the United States.

school choice, naming patterns, investment in public goods, or voluntary conscription—are either unavailable or exhibit little variation in our setting.[6] Instead, we exploit the coincidental existence of an exceptional repository of photographs covering our settlements, and material available due to the grassroots nature of community building in our environment: organizational meeting records, local inter-ethnic pacts, and the local rise of societal leaders. Pre-state Jewish nation-building combined top-down ideology with decentralized, community-led governance: while there was a central ideology of melting pot and the in-gathering and integration of the "exiles" (Horowitz et al., 1978; Gorny, 2001), the Israeli polity was also characterized by a mosaic of institutional sub-centers with quasi-governmental functions. Our findings highlight an important, novel role for these local inclusive institutions and democratic structures (Sellers et al., 2020), through the local production of societal leaders.

Our study also speaks to a literature that examines the role of leaders in: influencing whether and how strongly individuals identify with their nation, sometimes with negative consequences (Assouad, 2020; Cagé et al., 2023); shaping socio-economic outcomes via their personal networks (Bai et al., 2022); or mobilizing masses and fostering coordination in decentralized settings (Dippel and Heblich, 2021). A crucial difference between those studies and ours is that we consider the rise of leaders as an *outcome* of the local institutional framework and aim at isolating exogenous variation in the upbringing conditions of possible leaders within a nascent, informal state.

Our research is motivated by a literature on ethnic fragmentation and the functioning of the state. Migrants arriving to Palestine under the British Mandate managed to create a unified national identity in a relatively short period of time—further cemented by the establishment of the state of Israel in 1948 and the victory in the First Israeli-Arab War (1947–1949). The existing literature finds ambiguous effects of diversity: social divisions might persist (Dippel, 2014) and population heterogeneity has been associated with weak public good provision and conflict (Michalopoulos and Papaioannou, 2016; Miguel and Gugerty, 2005); intergroup contact in multi-ethnic societies has however been shown to be conducive of cooperation (Allport, 1954; Lazear, 1999; Bazzi et al., 2019) and to shape political behaviors (Brown et al., 2021). In our context, we find that the interaction of diversity and strong local institutions might nurture social cohesion and the rise of efficient leadership and organization at the local level.

Our work builds upon the research looking at the peculiar institutions of Manda-

[6]The first settlement-level census data, in 1948, records Hebrew as the main language of 80% of the population, leaving little cross-sectional variation for analysis. Other conventional nation-building outcomes, such as local public good provision, cannot be systematically measured at the settlement level. Likewise, no systematic data exist on voluntary military conscription before 1948, and military service became compulsory after the establishment of Israel. The documented emergence of military *leaders* may, however, partly capture this dimension.

tory Palestine and the early state of Israel. Rayman (1981) and Rabin (2013) discuss local leadership in kibbutzim (integrated "utopia" where everyone has the desire to promote the welfare of the group, see Spiro, 1956) and how they contributed to nation-building. We do find that kibbutzim were instrumental in the organization of the state of Israel through the nurturing of leaders rather than through the scaling-up of their own societal model. The limits of their egalitarian principles in an otherwise capitalist world are indeed discussed in Abramitzky (2008), Abramitzky and Lavy (2014), or Abramitzky (2018). Our work finally relates to the research looking at the conflict between Israel and Palestine, sometimes studying early tensions during the British Mandate (Panza and Swee, 2023). Our contribution is to look at the assimilation of Jewish and Arab settlements through the analysis of meetings in Jewish settlements and historical records from the First Israeli-Arab War.

The remainder of the paper is organized as follows. Section 1 presents the historical context. Section 2 and Section 3 describe our data sources and the empirical strategy. Section 4 establishes the baseline results and discusses robustness checks. Section 5 focuses on mechanisms. Finally, Section 6 briefly concludes.

1 Background

1.1 The Jewish settlements of Mandatory Palestine

The presence of Jewish settlements in Palestine steadily grew during the British Mandate. Land colonization followed a strategy shaped by the logic of frontier expansion, mostly aimed at increasing the size of existing settlements and ensuring their economic stabilization.[7] From the 1930s, the main Zionist settlement agencies—the Jewish National Fund and the Jewish Agency—focused on adding land to established settlements with the goal of creating large and continuous blocks in the coastal plain and beyond (Penslar, 2000, p.216). Several principles guided this territorial expansion strategy: the targeting of key geopolitical locations to define the borders of the future state of Israel (Near, 1992; Abramitzky, 2018); the expansion into areas with minimal or no existing Jewish presence, including lands near Arab towns and villages; and the strengthening of such areas through new settlements in their vicinity. A striking illustration is the "tower and stockade" strategy, implemented in response to the 1936–1939 riots and based on an Ottoman legal clause preventing the demolition of buildings once their roofs were completed. Over fifty agricultural settlements were hastily erected in the proximity of

[7] Zionist settlement practices explicitly drew on European colonial models of expansion, and land was often acquired through transactions with absentee landlords and third-party intermediaries, resulting in the displacement of Palestinian cultivators (Reichman and Hasson, 1984; Alff, 2023; Eldin, 2019).

existing settlements using this method, optimizing control over large areas with relatively few settlers—a general strategy that we will exploit in Section 3.

Settlement location was not only guided by strategic purposes, but also by economic considerations. Geographic, topographic, and hydrological surveys were often conducted prior to land purchases, with the objectives of (i) making informed recommendations about land suitability and (ii) estimating land improvement costs and the time necessary for settlement (Stein, 1984). In addition to agro-climatic conditions, particular attention was paid to drinking water availability, swamp drainage potential, and proximity to main roads and railways. Both strategic expansion at the fringes of existing settlements and (first- and second-nature) geography will prove instrumental in predicting settlement creation, as will be shown in Section 3.

Finally, Jewish agricultural colonization typically took the form of three main types of settlements: the kibbutz (collective), moshav (cooperative) and moshava (small village with privately owned land/farms).[8] The settlers' socio-economic background and ideologies could be predictive of the settlements that they would join (Weintraub et al., 1969): for example, socialist Zionists would tend to favor the kibbutz. This agricultural colonization was coupled with the expansion of urban quarters to strengthen Jewish presence in Palestine's towns, which hosted the majority of migrants (such expansion happened in Tel Aviv, Haifa, Jaffa, and Jerusalem, see Hasson and Gosenfield, 1980). While there were many newly created agricultural settlements (our objects of interest), relatively fewer immigrants settled in them.

1.2 Jewish migration to Mandatory Palestine

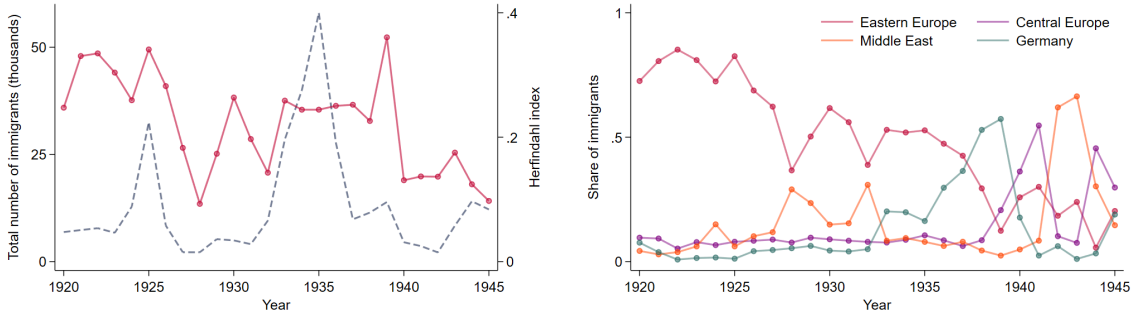
Between the late 19th century and the end of the British Mandate in 1948, waves of Jewish migration (*aliyah*) brought approximately 480,000 Jews to Palestine, more than 80 percent of whom were Ashkenazi Jews from Europe.

Early settlements of Palestine In 1800, the Jewish population of Ottoman Palestine consisted of 6,700 people, of which only 100 settled in rural areas (Bachi, 1974). The first *aliyah* (1881–1903) marked the beginning of Jewish agricultural settlements in Ottoman Palestine, with migrants primarily from Eastern Europe and Yemen. The second *aliyah* (1904–1914) brought to Palestine around 40,000 settlers, mostly from Russia and Poland, who pioneered new forms of collective agriculture centered on mixed farming and intensive cultivation: the *kvutza* (precursor to the kibbutz) and the *moshav*. These early

[8] Due to in-house military training, kibbutzim were more often located in insecure areas where ethnic relations could turn violent (Penslar, 2000). During years of relative peace between Jews and Arabs, moshavim were founded at greater rate than kibbutzim, while the opposite occurred when ethnic relations became more violent (Muchlinski, 2013).

efforts were often marked by limited experience, unfamiliarity with local environmental conditions, and a relative shortage of financial resources which lent it a trial-and-error character (Weintraub et al., 1969). The typical migrant was young, single, male, poor, influenced by socialist ideas, and imbued with a strong national ideology (Kimmerling, 2008). Prominent figures from this period—such as David Ben-Gurion, Berel Katzenelson, Yitzhak Ben-Zvi, and Yitzhak Tabenkin—later became key political leaders in the Jewish community of pre-state Palestine (*Yishuv*).

Figure 1. Jewish migration to Mandatory Palestine (1920–1945).



(a) Migration patterns

(b) Origins

Notes: Panel (a) displays the yearly number of immigrants into Palestine (blue, dashed line) and the concentration of these yearly migration flows across origins (red line, with a Herfindahl index computed using each country of origin). Panel (b) displays a subsample of their regions of origins: Eastern Europe, Central Europe, Middle-East, and Germany/Austria—Western Europe (excluding Germany/Austria), Maghreb and the rest of the World are not shown.

Jewish migration to Mandatory Palestine Our period of interest begins after 1919–1920 and the establishment of Mandatory Palestine. Jewish migration remained a contentious issue throughout the Mandate, as the Zionist movement continuously pressured the British authorities to permit increased immigration, framing it as the "in-gathering of exiles" (Hacohen, 2003). As shown in panel (a) of Figure 1, approximately 10,000 Jewish immigrants arrived annually in Mandatory Palestine between 1920 and 1945, with notable surges in 1925—after the United States passed the Immigration Act of 1924 thereby redirecting flows from Europe to Palestine—and around 1935, in response to rising antisemitism in Germany. These flows were initially dominated by migrants from Eastern Europe (Poland, Russia, Ukraine, Belarus), with increasing representation from Central Europe, Germany, Austria, the Middle East, and Western Europe after 1928.

Migration not only fluctuated in volume but also evolved in character, owing to time-varying conditions across typical origins. The third *aliyah* (1919–1923) brought around 37,000 predominantly low-income Jews from Russia, Poland, and Hungary, many of whom embraced Zionist ideology and contributed to the growth of the kibbutz move-

ment. The fourth *aliyah* (1924–1929) was largely composed of merchants and artisans migrating with their families, leading to the establishment of the first settlements specializing in citrus production—a sector which became the driving source of export revenue for Jewish agriculture. The fifth *aliyah* (1930–1939) marked a dramatic demographic shift, bringing nearly 250,000 immigrants and raising the Jewish population to approximately 450,000. Initially dominated by Polish pioneers, this wave increasingly included middle-class German Jews fleeing Nazi persecution, along with migrants from Romania, Bulgaria, Serbia, and Czechoslovakia. After 1939, immigration was hampered by the British White Paper of 1939 and a large part of it became illegal and clandestine (*aliyah bet*); a specific branch of the Haganah played a central role in organizing travel from Europe and the Middle East to the port of Haifa or the border between Palestine and Syria, where migrants were transferred to migration camps (Hacohen, 2003). Figure 1 illustrates the variation induced by these migration dynamics in terms of the origin of the typical migrant: a Herfindahl index summarizing the annual concentration of origin countries in panel (a); and the distribution of regional origins over time in panel (b). Interestingly, diversity in migrant origins fluctuated significantly over time, in a manner that is orthogonal to the general incidence of migration.

Importantly, changes in migrant *origin* were accompanied by changes in motivation and preparedness. Early migrants of the third *aliyah* were often driven by Zionist ideology and recruited through youth movements like "HeHalutz", which organized training programs and promoted communal settlement models (Near, 1992). By contrast, later waves of migration were both far more numerous and driven by strong, exogenous factors at origin—such as discrimination and persecution. These later arrivals tended to be less politically motivated, less prepared for agricultural life, and less "pro-active" compared to earlier settlers. Many of those willing to join rural communities would rely on the Jewish Agency's Rural and Suburban Settlement Company to direct them toward suitable, strategic areas of settlement (Trezib and Sonder, 2019).

1.3 Jewish identity in the Yishuv

A core objective of Zionism was the creation of a common national consciousness and the transformation of Jewish culture. Indeed, if Jewish immigrants to Palestine differed in religious belief and practice (Goldberg, 1985), they were united by a central historical narrative: that of a "people without land" (Horowitz and Lissak, 2012). The formation of Jewish socio-political institutions under British rule was entwined with the creation of a new national identity for the Jewish people.

The Jewish population in Mandatory Palestine was remarkably diverse. Although Mizrahi Jews from Yemen and Iraq arrived during the Mandate, Ashkenazi Jews were the

dominant group. Yet, even among Ashkenazim, there were distinct subgroups: German-speaking *Yekkes* were highly assimilated into German culture; *Oberlanders* from the Habsburg Empire adopted German or Hungarian as their first language; *Unterlanders* from the northeastern regions of the historical Kingdom of Hungary predominantly spoke Middle Yiddish; *Litvaks* from present-day Lithuania, Latvia, and Belarus spoke Eastern Yiddish; and Galician Jews from western Ukraine and southeastern Poland primarily spoke Yiddish. Despite ethnic, linguistic, and cultural commonalities (e.g., among Jews from the Pale of Settlement, Poland, and the Baltic states), their pre-migration experiences shaped their cultural and ideological orientations in varied ways (Sternhell, 2009).[9]

Although migrants shared strong cultural ties, differences in language, culture, ideology, values and preferences posed significant challenges to nation-building. The challenges induced by such diversity in the formation of a national identity have been extensively documented in the literature (Weingrod, 1985; Gorny, 2001; Smooha, 2018; Ya'ar, 2005; Yitzhaki and Schechtman, 2009), highlighting the assimilation issues of Mizrahis (Weingrod, 1985), Yekkes (Sela-Sheffy, 2006; Kranz, 2016), or Jews from Bulgaria speaking Ladino/Judaeo-Spanish (Ben-Bassat, 2016). For instance, the divide between Sephardim and Ashkenazim—differing in lifestyle, occupation, wealth, but also religious practices—culminated in the establishment of a dual Rabbinate in 1917.

The diverse traditions imported by immigrants made the *Yishuv* a social laboratory. Zionist nationalism required immigrants to forgo their own exilic identities in favor of a new, collective settler society—a process often described as the "blending of the exiles" (Kimmerling, 2008). This melting-pot approach, which became an official policy after the establishment of the state of Israel (Yitzhaki and Schechtman, 2009), originated in the Mandate era, when models of assimilation were shaped by ideals of the "good society" and the "new Jew" (Ya'ar, 2005). Cultural production played a central role in this transformation, mostly through language or education, and even the historical Jewish population of Palestine started to adopt the new cultural practices (Rozin, 2018).

1.4 The 1947–1949 War

The settlement of Jewish migrants in Mandatory Palestine was accompanied by land disputes, riots, and violence between Jews and Arabs; inter-ethnic violence escalated into a civil war in 1947–1949, the declaration of independence of the state of Israel, and

[9] Countries and regions of provenance would have generated dividing lines between migrants. However, most settlers were willing to forgo their previous ways of life, and they had overlapping values and objectives prior to joining settlements of Mandatory Palestine. For a minority of settlers, pre-settlement ties were *very* strong: the movements "Hashomer Hatzair" or "HeHalutz" were organizing and training groups in Europe (Yona, 2012), and many early pioneers forming new settlements had training in such farms where they would have built connections.

the *Nakba*—the expulsion of 700,000 Palestinian Arabs from their homes (Khalidi, 1997).

The civil war started after the United Nations resolution to end the British Mandate over Palestine—an objective already anticipated in the 1939 White Paper. The fighting was short and brutal. Until early 1948, neither side gained the upper hand; thereafter, Jewish forces unleashed a series of major attacks, prompted by the impending withdrawal of the British Mandatory powers and the prospect of an invasion by the neighboring Arab states (Morris, 2004). These operations were formalized in Plan Dalet, designed to secure territorial contiguity for the emergent Jewish state. Most operations were carried out centrally by mobile regional infantry brigades (e.g., 317 villages were occupied by "national forces" versus 142 by local forces, see Morris, 2004), while local and non-local Arab fighters were organized under the Arab Liberation Army. Brigade commanders were given lists of villages to target, but the guidelines allowed some discretion within their zone of operations (Morris, 2004): some communities were more peacefully evacuated; some expulsions involved the killing of non-combatants after they had surrendered. Interpretations of these events remain contested: while mainstream Zionist accounts describe a voluntary flight under instructions from Arab leaders (Gelber, 2006), Arab historians and revisionist scholarship attribute the exodus primarily to Zionist military operations and policies, framing it as integral to the creation of the state of Israel (Khalidi, 1988; Masalha, 1992; Pappe, 2015).

Local communities played a role during the war. Local leaders in locations where Jewish and Arab villages expected repeated interactions—places designated as border areas by the UN partition plan—were focusing on peace-building (a behavior discussed as a rational response to wartime dynamics and uncertainty in Haran Diman, 2024). These agreements (documented in Azoulay, 2014, p.414) were mutual initiatives by neighboring Jewish and Palestinian communities to resist the violence led by their respective national movements. Sometimes, the killing of civilians was partly avoided thanks to the intervention of inhabitants from nearby Jewish settlements who informed Arab villages of an incoming attack or directly interfered to stop soldiers.[10]

The Palestine bill of 29 April 1948 formalized the end of the British administration over Palestine, closely followed by a "Declaration of the Establishment of the State of Israel". Our period of interest does not include the early years of Israel, but we rely on key figures of the state after the war in order to identify the nurturing role of settlements.

[10]For example, nearby settlers convinced the military not to intervene and spare the Arab population in the villages of Furaydis and Jisr a-Zarqa. The peaceful coexistence of Jewish and Arab communities before the war did not provide a guarantee against depopulation, given the centralized nature of most military operations, e.g.: the village of al-Ja'una was depopulated during Operation Yiftach, despite having had good relationships with its Jewish neighbors in Rosh Pina; the village of Deir Yassin saw the killing of more than a hundred Arab civilians by paramilitaries (Hogan, 2000), despite having signed a non-belligerency pact with its Jewish neighbors.

2 Data

This section describes our primary data sources; we leave the more detailed, comprehensive presentation of our data to Appendix A.1.

Settlement characteristics We digitize all available censuses during our period of interest. These data sources include: village- and settlement-level population data from the 1922 and 1931 censuses of Palestine; information collected for the Palestine Index to Villages and Settlements (1944–1945); the 1945 Village Statistics; and later Population Censuses (1945–1948). The latter two are our main data sources which contain information at the settlement level on their date of creation, their composition (total population, regions of origin, language, occupations, industries), and their agricultural activity. We use these data to define our endogenous measure of settlement diversity: a Herfindahl index computed across 8 large regions of origin chosen by surveyors at that time (e.g., "Eastern Europe") and reflecting population composition around 1945. These surveys also report information collected during previous waves such that we observe populations at different points in time for early settlements. Note that we do not fully exploit earlier censuses, because our baseline empirical strategy hinges on variation induced by the exact timing of settlement creation, and earlier censuses would then only allow us to exploit (less than) 10 years of settlement creation. We extract the following further data on the organization of Jewish settlements from two "censuses" of Jewish Agriculture in 1927 and in 1936 and two "industrial censuses" in 1929–1930 and in 1936: investment by the Jewish Agency, agricultural output by crops, and occupations at the 3-digit level.

Migration We capture the timing and nature of migrant inflows with aggregate statistics based on the Card catalog of the Statistics Bureau of the Jewish Agency's Immigration Department, 1919–1948. More specifically, we digitize the annual number of Jews admitted into (or refused entry into) Palestine and their characteristics, for instance, their financial resources, their reported occupations, and their country of origin.[11] The "shift" in our shift-share empirical design will collapse the characteristics of new immigrants

[11]We compile the number of Jews admitted into (or refused entry into) the United States, as the outside option for Jews willing to leave their current country of residence to "new countries"—we also compile migration to South America, Canada, or South Africa. Over the course of the British mandate in Palestine, as many Jews migrated from Europe to Palestine as to the United States—albeit with a different timing. Indeed, while migration to the United States dried up markedly after 1925 (following The Immigration Act of 1924) and from 1930 (following the Great Depression), it increased again from 1938–1939 when migration to Palestine became more restricted—as induced by the White Paper of 1939. We provide evidence about the substitution across destinations in Appendix A.1, together with a discussion of non-Jewish migration to Mandatory Palestine and Jewish *emigration* from Palestine.

into a yearly Herfindahl index computed using their *country* of origin (see Figure 1).[12]

Measurement of community relationships Given the scarcity of data characterizing social relations (i) within Jewish communities and (ii) with Arab villages, we rely on an array of diverse, imperfect data sources: meeting notes, photographs—both prior to the war—and historical reports on the 1947–1949 war.[13]

First, we exploit historical reports on inter-community relationships by coding the information contained in Azoulay (2014) and collected from the Haganah archives about civil pacts between Arab and Jewish communities (see Appendix A.1), and by digitizing and georeferencing a map on the destruction of Arab and Jewish localities. The former dataset provides information about 100 village pacts prior to or during the conflict, with a description of their nature (e.g., peace agreements, warnings of a forthcoming attack). The latter data have been combined with details on the village depopulation date, the type of Jewish attack, the presence of non-combatant casualties, and whether the local militias defended the village, drawing on Abu-Sitta (2000) and Khalidi (2008). These sources indicate that a total of 448 villages were depopulated, 69 of which witnessed the killing of unarmed civilians. In our regression sample of 316 Jewish settlements, we observe that the probability of being located within a 10km radius of an Arab village with civilian casualties is 72%. Depopulation was more widespread and homogenous within sub-districts such that there will be limited empirical variation for this specific outcome.

Second, our analysis of within-community cohesion and inter-community relations draws on a systematic collection of meeting minutes from various Jewish settlements in Mandatory Palestine. The sources are the Yad Tabenkin archives (covering kibbutzim affiliated with the United Kibbutz Movement, HaKibbutz HaMeuhad, and Chever HaKvutzot) and the National Library of Israel (covering the moshavim). These archives document hundreds of meetings held between 1930 and 1947, primarily in the following

[12] In our baseline shift-share instrument, we treat immigrants from Poland differently from those originating from Soviet Ukraine, and equally so as Mizrahi Jews from Yemen. We thus do not weigh different groups by country-pair similarity in language or culture. However, our *endogenous* Herfindahl index—based on actual settlement composition—does hinge on larger regions, as grouped by the surveyors administering the village survey; those larger categories were possibly more relevant markers of identity to the Jewish population. In robustness checks, we use instruments constructed from rougher categories, and we control for the separate prevalence of each region of origin within settlements to ensure that our findings are not driven by certain migrant origins.

[13] These data sources are imperfect, in that they partially cover our baseline sample of settlements and might be subject to selection in their production and conservation. For example, meeting notes and photographs are likely to be produced (and preserved) in more established and better-organized communities, and we find indeed that their geography, composition, and organization differ from the average settlement: Appendix B.6 provides a systematic analysis of differences across all settlements, kibbutzim, settlements covered by meeting notes, and settlements covered by photographs. A corollary of such selection is that, while meeting notes and photographs might provide a varied and subtle account of community building, their limited coverage prevents us from applying our baseline empirical strategy.

locations: Alonim, Ashdot Yaakov, Beit Hashita, Beit Oren, Dalhamia/Gesher, Degania A, Degania B, Ein Gev, Ein Harod, Ein Hayam, Ganigar, Gedera, Geva, Givat Brenner, Givat Hashlosha, Hadera, Haifa, Hulda, Kfar Saba, Na'an, Ramat David, Revivim, Shafayim, and Tel Aviv. In total, these resources cover more than 6,000 pages of meeting transcripts; the material is unstructured and not always of direct interest: meetings sometimes discuss concrete issues arising in certain settlements; they also cover general issues, organizational issues, or the ideology of the movement. We digitize, transcribe, and process these materials using a combination of OCR, manual transcription (for handwritten notes), systematic correction, translation, and large language model-based topic analysis.[14] From these meeting notes, we extract 4,732 entries, each tagged with the settlement, discussion topic (e.g., immigration, cohesion, economy, education, land, language, security), tone (e.g., critical, urgent, constructive), an importance score (0–100), and illustrative quotes. Notably, meetings often address issues across multiple settlements and include representatives from different communities. As a result, although meetings were concentrated in 30–40 locations, our analysis identifies 107 distinct settlements with extractable topics/issues. To study *inter*-community relations from the Jewish viewpoint, we develop a parallel procedure, isolate 270 discussions of Arab neighbors, and classify these entries across their location, topic, tone, and a positivity score from 0 to 100. We provide a more exhaustive description of these minutes in Appendix A.2.

Third, we retrieve the Eric and Edith Matson Photograph Collection from the online catalog of the Library of Congress; this collection of more than 20,000 glass plate negatives in reasonably good condition mostly covers Mandatory Palestine between 1918 and 1948 and were produced within the American Colony Photo Department located in Jerusalem. Each picture comes with a detailed description of its subject, location, and date, as partly curated by the Library of Congress. A caveat of the data sources is that we do not observe the name of the photographer or the reason why the picture was taken. We run a textual analysis of the title and description of subjects, as well as an image recognition algorithm to extract the objects depicted by each picture. This process allows us to classify pictures as aerial photographs, landscapes, daily scenes, etc.; of those, we select 3,385 pictures which can be assigned to one of our settlements, excluding Jerusalem and Tel Aviv to reduce the computational cost. We use a large lan-

[14]We correct and translate the meeting minutes using two technologies: DeepL Translate; and a large language model (GPT-4o). The main reason for using two parallel algorithms is to provide a validation of the output of the large language model, which would help discipline its hallucinations. Along the same lines, we consider three parallel procedures for the topic analysis: an unsupervised procedure based on BERT models developed and adapted to Hebrew by Chriqui and Yahav (2022) and Shalumov and Haskey (2023); a transformer-based topic modeling based on BERT and applied to the translated minutes (BERTopic, see Grootendorst, 2022); and a large language model (GPT-4o) whose full API-based prompt can be found in Appendix A.2 for the sake of transparency. In practice, the output from the externally-prompted large language model proves better than these alternative "in-house" options.

guage model accommodating both a textual and image input (GPT-4o) to classify (a) the emotions conveyed by each photograph (anger, amusement, awe, contentment, disgust, excitement, fear, sadness) and, when relevant, (b) the nature of social relationships conveyed by the picture (hostile, amicable, neutral). The output of this procedure, better described in Appendix A.3, is a location, a date, a picture type, a list of depicted objects (e.g., "two women, mill, grinding stone, building, door, ladder, stones, bowl"), a vector of probabilistic emotions and a vector of probabilistic social relationships.

Societal leaders Our main focus is on individuals who went on to play visible and lasting roles in shaping the Israeli state *from 1948 onward*—whether through political engagement, military responsibilities, or cultural contributions.[15] We refer to these individuals as *societal leaders*. Guided by this definition, we compile detailed biographies of about 250 *major* political, military, and cultural leaders who participated in Israel's nation-building after 1948 and who had lived in a Jewish settlement of Mandatory Palestine before 1948.[16] Specifically, we collect information on year of migration, type and duration of government or military role if any (Knesset member, president, minister, Haganah commander, etc.), country of birth, places of residence in Palestine, publications, and political affiliation if any.[17] Within our sample of 316 settlements,roughly 70 were home to at least one future leader during the Mandate. Finally, we analyze the political activities of these leaders through a systematic examination of their speeches and interventions in the Knesset, focusing on the subsample of approximately 120 individuals who served as Knesset members or ministers between 1949 and 1977.[18]

Two issues arise with the previous selection of leaders (which we list in Appendix C).

[15]This definition includes notable writers (novelists, essayists, poets, playwrights), composers and songwriters, film directors, and journalists or editors.

[16]This data collection has benefited from the outstanding research assistance of Susan Staszewski. Note that a few of these key leaders already had an important role during the Mandate: David Ben-Gurion, the first prime minister of Israel, born in Plonsk (Poland), who migrated to Palestine in 1906, and spent most of his life during the Mandate in Tel-Aviv where he led the Jewish Agency (from 1935); Yitzhak Tabenkin, born in Bobruisk (modern-day Belarus), who migrated to Palestine in 1912, established the first kibbutz proper in Ein Harod, and was elected Knesset member in 1949–1951 and 1955–1958; Moshe Dayan, born in Kibbutz Degania (Tiberias) in 1915, who was one of the key military figures of the Yishuv and had a ministerial career covering Agriculture (1959–1964), Defense (1967–1974) and Foreign Affairs (1977–1979). We will run a sensitivity analysis excluding pioneers with political responsibilities *during* the Mandate.

[17]We rely on the following websites to select and compile our dataset: the Knesset (the unicameral legislature of Israel), Israel Defense Forces, the Jewish Virtual Library, and the Jewish Agency, all of which provide short biographies for each leader. We complement them with secondary sources, including biographies and autobiographies for a selection of individuals, as detailed in Appendix C.

[18]This analysis draws on around 3,340 minutes of Knesset sessions covering: 1st Knesset (1949–1951); 2nd Knesset (1951–1955); 3rd Knesset (1955–1959); 4th Knesset (1959–1961); 5th Knesset (1961–1965); 6th Knesset (1965–1969); 7th Knesset (1969–1974); 8th Knesset (1974–1977). We process these records using a Large Language Model to extract: (i) the main topics of discussion, the speakers involved, and the tone of their contributions; and (ii) specific references to Arab-Jewish cohabitation. Details on data collection, processing, and description are provided in Appendix A.4.

First, it involves some discretion on who to include. We thus rely on the 19 volumes of the *Encyclopedia of the Founders and Builders of Israel* curated by Tidhar (1947) to provide a robustness check on a larger, external sample of about 6,000 influential settlers for whom we have less extensive information. Second, several leaders resided in different locations, hence requiring a baseline allocation mechanism. The process of settlement allocation is straightforward for the subset of leaders who resided in the same location throughout the Mandate (around 60% of them). For those who lived in multiple locations, we identify the first place of proper, perennial residence upon arrival to Mandatory Palestine—dropping temporary locations unlikely to have played a major role in shaping their experience within the Yishuv. For individuals who moved to a different location during early childhood, we assign the location where they grew up during their "impressionable years". To reduce concerns that our findings depend on these choices, we consider alternative allocation procedures, e.g., using probabilistic weights or removing mobile leaders, in sensitivity checks.

Small settlements played a disproportionate role in the fostering of societal leaders. A snapshot of our dataset shows that Givat Hayim (a kibbutz founded in 1932) hosted: Baruch Azanja, a lawyer who migrated from Danzig in 1933 and became a member of the Knesset for Labor parties; Yitzhak Ben-Aharon, general secretary of the Histadrut and born in the Austro-Hungarian empire; German-born couple Giora Yoseftal, a minister who worked actively in the Jewish Agency as head of the absorption section, and Senetta Yoseftal. Ein Harod (a "precursor" kibbutz founded in 1921) hosted: Warsaw-born Katriel Katz, Hanagah commander, founder of the *aliyah* camps and later diplomat; Haganah commander Hadar Kimchy; Yitzhak Tabenkin (after which the Yad Tabenkin was named), an early member of the Knesset born in Bobruisk; and Russian-born novelist and editor Avraham Shlonsky. Rehovoth and Kfar Marmorek hosted: Chaim Weizmann, the first president of the state of Israel; writer and politician Yizhar Smilansky; Moshe Smilansky, writer and Zionist leader advocating for a peaceful coexistence with the Arabs; or Ya'akov Fichman, poet and winner of the Israel Prize for Fine Literature.

These selected examples illustrate the following empirical regularities that will underlie our analysis. First, many key figures of the early state of Israel grew up and lived in small agricultural settlements of Mandatory Palestine (a third of them in our sample), especially kibbutzim (a well-known fact discussed in Rayman, 1981). Second, these settlements were sometimes diverse, both across the origins of settlers and their future political ideologies. Third, the establishment of settlers in certain locations was often tied to their date of immigration and to the creation of the settlement itself, but not always. Sometimes, migrants would move from their initial locations to other places, either for economic or ideological reasons. Sometimes, they would join a long-established

settlement. For these reasons, the composition of settlements is not "exogenous" and might depend, among many other factors, on the identity and charisma of its founders. To address this issue, we develop the following shift-share design.

3 Empirical strategy

3.1 A shift-share design

Agricultural settlements were created at different points in time and populated by immigrants of different origins, wealth and skills. The settlement process was possibly disciplined by strategic considerations from the Jewish Agency, by an expansion of the frontier around existing settlements and by the individual location choices of migrants. In what follows, we develop a procedure to predict the dynamics of settlement creation from (i) geographic conditions, (ii) the initial allocation of the Jewish population and (iii) a systematic, spatial expansion of the frontier. We then describe how we combine such a dynamic prediction (a "share" across possible locations in Mandatory Palestine) with the annual composition of migrant inflows (a "shift") to predict the allocation of settlers with different characteristics across space.

Suitability and initial settlements There exist various geographic factors which explain the allocation of Jewish settlements across space. Agro-climatic conditions are challenging in this region with high temperature, water scarcity and a soil that is not suitable to many staple crops grown in Europe. We divide the territory covered by Palestine under the British mandate in hexagons of about 2-kilometer radius and we compute for each of these geographic units: latitude and longitude; maximum, minimum and average elevation; slope; soil bulk density (Hengl, 2018); organic carbon content (Hengl and Wheeler, 2018); the average annual precipitation and temperature (Hijmans et al., 2005); the distance to the nearest shore, river, canal, stream, wadi; the length and area of water bodies; the distance to and length of roads and railways as of 1920; the potential yield per hectare for various crops (citrus in particular); the distance/travel time to the nearest existing city, to Haifa, to Jerusalem and to Tel-Aviv/Jaffa; the density, distance/travel time to existing Arab settlements; the share of land claimed by the Jewish companies and private owners and by the Jewish National Fund; and later land regulations imposed by the British mandate. We use all these controls, further described in Appendix A.1, to levy variation in the probability to host a Jewish settlement at the end of the period, in 1947. Letting i denote a given hexagon, we regress the probability to have at least one Jewish settlement as of 1947, $p_i \in \{0, 1\}$, on the above-described set of controls, $\mathbf{X}_i$, and we extract a uni-dimensional predictor, $\widehat{p}_i$. We will refer to this object as the "suitability"

to host a settlement and combine it with the initial distribution of settlements in 1920 to predict the *timing* of their future allocation across space.

Figure 2. Settlements in Mandate Palestine, suitability and initial allocation.

(a) Settlement suitability (b) Initial allocation

Notes: Panels (a) and (b) respectively show the distribution of settlements in 1947 and their initial allocation in 1920 across Mandatory Palestine. The main geographic unit is a hexagon with a radius of 2 kilometers. In panel (a), colors indicate the probability to have a settlement in 1947, as predicted from the set of controls, $\mathbf{X}_i$ (colors indicate quantiles: low in green, intermediate in yellow, and high in red). The estimates are reported in Appendix A.5; the geography of Mandatory Palestine is shown in Appendix Figure A3.

Figure 2 displays the predicted suitability to host settlements, their actual distribution in 1947, and their initial allocation in 1920 across Mandatory Palestine. A large region along the coast (a "citrus belt") and extending into the hinterlands towards Jerusalem, Nazareth and the North of the country, is deemed suitable to host Jewish settlements. Actual settlements are not uniformly allocated within this suitable region; they are instead clustered along a few axes. This selective allocation of settlements can be better understood by looking at the initial allocation of settlements: the later communities appear to connect these earlier settlements. The next section relies on this insight—an expansion of the frontier at the fringe of existing settlements and in locations with high "fundamentals"—to predict the dynamics of settlement creation.

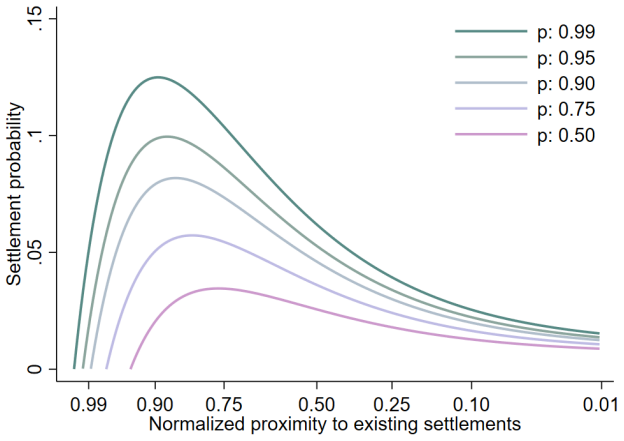
Frontier expansion We predict the dynamics of settlement creation as follows. To better account for the possible time lag between migration and settlement into agricultural communities, we first divide the time period between 1920 and 1947 into 3-year intervals and let $\tau = 0, ..., 9$ denote the bounds of these intervals. The objective is to

construct a mapping $f : \mathbf{s} \mapsto f(\mathbf{s})$ from the existing settlements, $\mathbf{s}$, to settlements in the following period, $\mathbf{s}'$, that is informed by agro-climatic conditions and other relevant measures of suitability for economic activity. In practice, we consider a measure of local settlement density, $s_{i\tau}$, which we define as follows for any location i in any given period τ characterized by a set of existing settlements I_τ:

$$s_{i\tau} = \sigma_0 \sum_{j \in I_\tau} \frac{1}{t_{ij}}, \tag{1}$$

where: t_{ij} is the travel time between location i and settlement j using the more exogenous, initial network of roads, railways, and waterways in 1920; and the parameter σ_0 is set such that $\{s_{i\tau}\}_{i,\tau}$ ranges between 0 and 1. This local settlement density ($\mathbf{s}$) is a market access measure, accounting for the spillovers exerted by each Jewish settlement.[19]

Figure 3. The (estimated) dynamics of colonization (1920–1947).



Notes: This Figure illustrates the nature of the mapping $f : (\mathbf{s}, \mathbf{p}) \mapsto f(\mathbf{s}, \mathbf{p})$. More specifically, we display the estimated probability of being in the proximity of a new settlement between $\tau - 1$ and τ (y-axis), as a function of the proximity to existing settlements in $\tau - 1$ (x-axis) for 5 different locations: a location j with a 0.99 percentile value in fixed suitability to host a new settlement, $\widehat{p}_i$, as induced by the set of observables; and locations with 0.95, 0.90, 0.75, 0.50 percentiles (darker green to purple). We report the estimates underlying this Figure in Appendix A.5 and Appendix Table A3.

Second, we identify the mapping f by estimating the following specification across locations and periods,

$$s_{i\tau} = \alpha_0 + \alpha_1 s_{i\tau-1} + \alpha_2 s_{i\tau-1}^2 + \gamma_0 \widehat{p}_i + \gamma_1 \widehat{p}_i s_{i\tau-1} + \gamma_2 \widehat{p}_i s_{i\tau-1}^2 + \varepsilon_{i\tau}, \tag{2}$$

where $\widehat{p}_i$ is the previous index of geographic suitability. In this estimated mapping, we include linear and quadratic terms in past settlement density to capture (i) the general

[19] We consider an elasticity to travel time equal to -1 in Equation (1) to mirror the standard market access measures. In any case, we cannot estimate nor calibrate this elasticity, as we do not observe measures of trade or social exchanges between locations across space.

expansion of the frontier at the fringe of *existing* settlements and (ii) the lower, possibly negative, returns to very high density. Indeed, settlements generally appear at a distance from an existing cluster of settlements—in a similar manner to train stations being set up at a given distance of existing stations. We allow these dynamic factors to interact with the pre-defined suitability, $\widehat{p}_i$, as settlements are set up within a subset of desirable locations. This specification is the minimal one capturing the dynamics of colonization: its spatial spread away from existing outposts, interacted with its dependence on geographic suitability. We illustrate the interaction underlying the (estimated) dynamics of colonization in Figure 3: the estimated probability of being in the proximity of a *new* settlement between $\tau - 1$ and τ is small for locations already in the proximity to *existing* settlements in $\tau - 1$. This probability then sharply rises and peaks between the 90th and 75th percentile in the normalized proximity to previous settlements; it then gradually decreases with distance to the frontier. This bell-shaped relationship is observed across locations with different initial geographic suitability ($\widehat{p}_i$); better suited locations however display a much sharper initial rise. In summary, colonization spreads on suitable land and in (not too) close proximity to existing outposts (with "direct visual communication with a neighboring settlement", see Near, 1992).

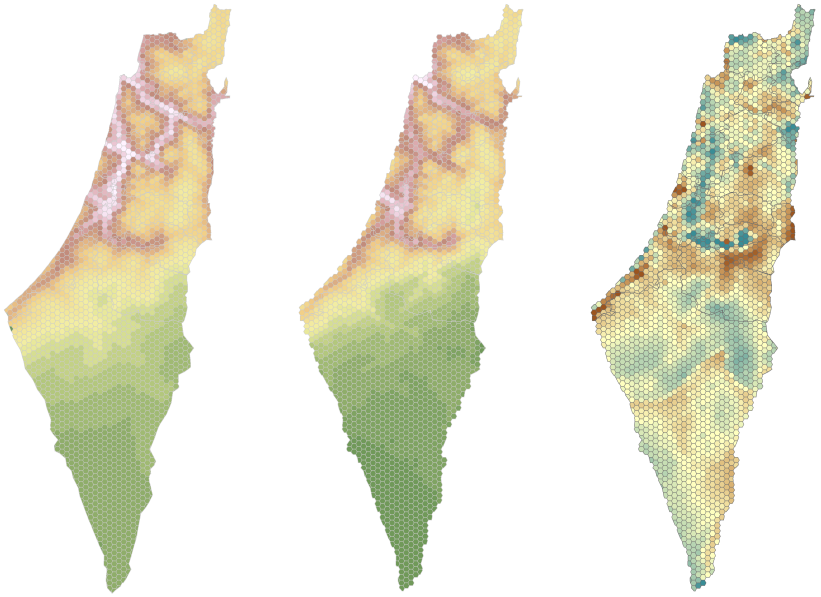
Third, we construct a dynamic prediction based on the initial distribution of settlements in 1920 and the previous forward projection.[20] The resulting predictions are orthogonal to the time-varying shocks affecting the settlement process between 1920 and 1947 and only relies on the timeless, *systematic* expansion of the frontier at the fringe of existing settlements and within suitable geographic areas. We represent the final predictions computed for $\tau = 1947$ in Figure 4: panel (a) displays the actual long difference in settlement density, $s_{i,1947} - s_{i,1920}$, between 1920 and 1947; and panel (b) displays the predicted long difference in settlement density, $\widehat{s}_{i,1947} - \widehat{s}_{i,1920}$, between 1920 and 1947. The two main ingredients for our prediction—geographic suitability and dynamic colonization at the fringe of existing settlements—allow us to match quite precisely the actual distribution of settlement creation between 1920 and 1947.

A shift-share design Our objective is however more ambitious than just fitting the *overall* creation of settlements: our design arguably captures the *timing* of settlement creation from 1920 to 1947 for any given location. We illustrate its predictive power by estimating the following equation, conditioning on fixed unobserved location characteristics and the overall timing of settlement creation in Palestine:

$$\Delta s_{i\tau} = \theta \Delta \widehat{s}_{i\tau} + \mu_i + \nu_\tau + \varepsilon_{i\tau}, \tag{3}$$

[20] We proceed by iteration, set the initial values as $\widehat{s}_{i0} = s_i$, the earlier settlement density in 1920, and define, $\widehat{s}_{i\tau} = \alpha_0 + \alpha_1 \widehat{s}_{i\tau-1} + \alpha_2 \widehat{s}^2_{i\tau-1} + \gamma_0 \widehat{p}_i + \gamma_1 \widehat{p}_i \widehat{s}_{i\tau-1} + \gamma_2 \widehat{p}_i \widehat{s}^2_{i\tau-1}$.

Figure 4. Predicting settlement creation and composition in Mandate Palestine (1920–1947).



(a) Actual density ($\mathbf{s}$) (b) Predicted density ($\widehat{\mathbf{s}}$) (c) Uniformity ($\widetilde{\vartheta}_i$)

Notes: The main geographic unit is a hexagon with a radius of 2 kilometers. Panel (a) and panel (b) show the long differences in actual and predicted settlement densities between 1920 and 1947 across Mandatory Palestine (ranging from 0 in dark green to 1 in white, while yellow then red indicates intermediate values). Panel (c) shows the variation in the (residualized) predicted measure of uniformity from brown (lower than -2.5 standard deviations) to yellow to blue (higher than 2.5 standard deviations) across Mandatory Palestine. The measure is a residual, cleaned from geographic variables. The variation arises from the interaction between the time-varying probability to host a new settlement and the time-varying composition of migrant inflows.

Table 1. Predicting the timing of settlement creation.

Change in settlement density ($\Delta s_{i\tau}$)	(1)	(2)	(3)
Predicted change in settlement density	0.652	1.037	0.797
	(0.075)	(0.073)	(0.042)
Observations	25,040	25,040	25,040

Notes: A unit of observation is a hexagon with a radius of 2 kilometers in a given time period (every three years between 1920–1947). All specifications include hexagon fixed effects and time fixed effects, and are specified in differences (for the dependent variable and the main explaining variable)—see Equation (3). In column (2), we control for hexagon-specific trends. In column (3), the explaining variable is the normalized variable $\widetilde{s}_{i\tau}$—thus providing a direct validation for the construction of the "shares" in the shift-share prediction (4). Appendix A.5 (Appendix Figure A13) provides a visualization of the estimated relationship reported in column (3). In these density-focused specifications at the hexagon-level, standard errors are computed following Müller and Watson (2022) to account for a flexible structure of spatial auto-correlation.

where μ_i and ν_τ represent hexagon and time fixed effects, respectively. We report the estimation of Equation (3) in Table 1. Our predictions very strongly correlate with the actual timing of settlement creation, even when we control for the secular, location-

specific acceleration of settlement creation over time.[21]

The dynamics of settlement creation, $\Delta\widehat{s}_{i\tau}$, as induced by geographic conditions, initial settlements and the systematic, spatial expansion of the frontier, are combined with the composition of migrant inflows to predict the allocation of immigrants with different characteristics across space. More specifically, consider θ_τ, a given characteristic of migrants arriving between $\tau-1$ and τ, e.g., the diversity of this migration wave across origins. We convert this characterization of the average *immigrant* over time into a characterization of *settlers* across space, $\widetilde{\vartheta}_i$. Letting $\widetilde{s}_{i\tau} = \Delta\widehat{s}_{i\tau} / \left(\sum_\tau \Delta\widehat{s}_{i\tau}\right)$, we have:

$$\widetilde{\vartheta}_i = \sum_\tau \widetilde{s}_{i\tau} \times \theta_\tau, \tag{4}$$

where the dynamic prediction, $\widetilde{s}_{i\tau}$, is the predicted "share" of settlers who arrived between period $\tau-1$ and τ in location i (whose predictive power is reported in column 3 of Table 1), and the composition of migrant inflows, θ_τ, is a "shift" over time. We provide an illustration of the variation induced by the shift-share design in panel (c) of Figure 4 where we display a measure of uniformity among *settlers*, $\widetilde{\vartheta}_i$, as predicted from the Herfindahl index of *immigrant* flows, θ_τ. Intuitively, we project the birthplace diversity of migrants across countries, $\{\theta_\tau\}_\tau$, onto the Palestinian territory to predict the cross-sectional diversity of settlers (ϑ) in each location.

A few important remarks are in order. First, our procedure hinges on an extensive list of controls to define settlement suitability, $\widehat{p}_i$. While our main objective is to extract variation in the *timing* of settlement, we do control for these separate elements in our baseline specification, and we do consider more parsimonious approaches in robustness checks. Second, the measure of uniformity among settlers, $\widetilde{\vartheta}_i$, is not a proper Herfindahl index, but a linearly-weighted average of Herfindahl indices. We do so in order to rely on recent advances in the econometrics of shift-share designs; we however provide a robustness check where the measure of uniformity is a non-linear construction from (linearly) predicted shares of each migrant group. Third, our shift-share design allocates migrants where settlement creation is likely, possibly before and after the settlement is actually created. Intuitively, we do not restrict migrant allocation to the exact date of settlement creation but to a "probabilistic period": migrants might join existing settlements, but they are more likely to be allocated to settlements that are likely to be created when they arrive (in relative terms, many migrants moved to existing *cities* throughout the period). While not confining settlement to one date, our procedure is however orthogonal

[21] In Appendix A.5, we provide a placebo test by including forward predictions to Equation (3). We also provide: further evidence of the dynamic process through which the settlement frontier expands to connect existing settlements, most notably maps of actual and predicted settlement creation over time; and a visual illustration of the first-stage relationships displayed in Table 1 (column 3) and Table 2 (column 2).

to the variation induced by the presence of charismatic leaders attracting already-located settlers from neighboring settlements.

Table 2. Predicting settlement composition across settlements.

Herfindahl index (ϑ_i)	(1)	(2)
Predicted uniformity ($\widetilde{\vartheta}_i$)	1.851	2.296
	(0.291)	(0.331)
Observations	316	316

Notes: A unit of observation is a settlement. The dependent variable is a Herfindahl index of origins across all settlers as measured in 1945; the explaining variable is a uniformity measure constructed from (i) the timing of settlement creation as predicted by the "frontier expansion" algorithm and (ii) a Herfindahl index of yearly migrant flows across countries of origin. All specifications include the variables used in the "frontier expansion" algorithm: all separate geographic characteristics used to capture suitability; and the initial allocation of settlements and travel time to settlements around 1920. We also include a trend and quadratic term in the year of establishment, as well as population-decile fixed effects. In column (2), we control for fixed effects at the level of sub-districts. Appendix A.5 (Appendix Figure A13) provides a visualization of the estimated relationship reported in column (2). Standard errors are clustered at the sub-district level and reported between parentheses.

We rely on such a shift-share design to isolate exogenous variation in the composition of settlers across settlements. Our shift-share prediction nested at the level of settlement strongly correlates with the actual Herfindahl index of origins across all settlers in 1945, ϑ. We show the relationship between these two quantities in Table 2, separately controlling for the elements entering the frontier expansion algorithm: the geographic characteristics used to capture suitability; and the initial allocation of settlements and travel time to settlements around 1920. We also include a trend and quadratic term in the year of establishment, as well as population-decile fixed effects. In column (2), we control for fixed effects at the level of sub-districts. The addition of these controls is important to support identification; we need the shift-share prediction $\widetilde{\vartheta}_j$ to be correlated with settlement-level outcomes only through its effect on settlement composition.[22]

3.2 Identification

Baseline specification Our baseline specification estimates the effect of settlement composition across locations j, ϑ_j, on our main community-building outcomes, y_j:

$$y_j = \alpha + \beta\vartheta_j + \gamma\mathbf{X}_j + \varepsilon_j, \tag{5}$$

[22]We report the respective distributions of ϑ_j and the shift-share prediction $\widetilde{\vartheta}_j$ in Appendix B.1, where we also show that our shift-share prediction is orthogonal to settlement type/organization. We also provide two validation exercises for our uniformity measure, $\widetilde{\vartheta}_i$, using: the allocation of newly-arrived migrants as recorded in the meeting minutes; and the allocation of influential settlers.

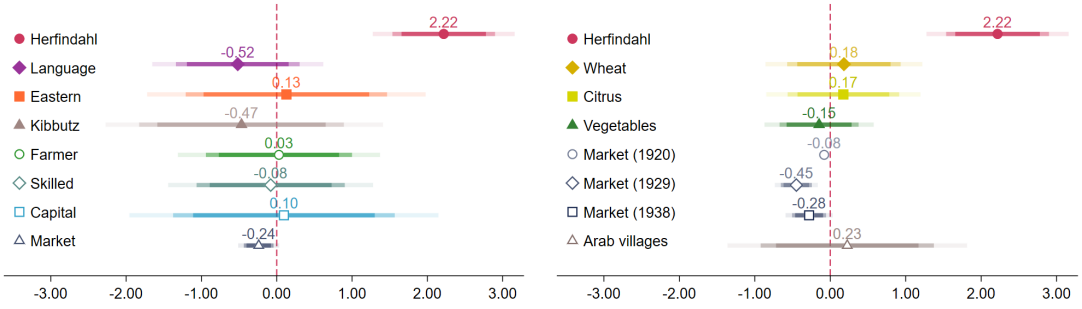
controlling for the various dimensions of settlement suitability, the travel time to initial settlements in 1920, sub-district fixed effects, population-decile fixed effects, and a quadratic function in the year of settlement creation ($\mathbf{X}_j$). Settlers' characteristics, ϑ_j, may however be endogenous to the dynamics of communities. For instance, the composition of new settlements could reflect strategic considerations from the Jewish Agency to facilitate the establishment of a Jewish state in Mandatory Palestine. Besides, the charisma of a few individuals may have affected the later location choices of migrants with specific characteristics. In the former case, the coefficient β would partly reflect omitted variation. In the latter case, it would be contaminated by reverse causality. To address these endogeneity concerns, we leverage the procedure developed in Section 3.1 and instrument settlement composition, ϑ_j, by our shift-share prediction $\widetilde{\vartheta}_j$.

Identification (in shift-share designs) Two assumptions underlie our identification: migrants are relatively more likely to settle in just-created settlements (*relevance*); and the predicted measure of uniformity is quasi-random with respect to the dynamics of communities (*exogeneity*). The identification of specification (5) relies on a recent literature discussing the latter assumption when applied to shift-share designs. Goldsmith-Pinkham et al. (2020) show that specification (5) provides consistent estimates if the shares, $\widetilde{s}_{j\tau}$, are numerous, dispersed and quasi-random. In spite of our efforts, this condition might not necessarily be satisfied: our shares are numerous and dispersed, but they might not be random with respect to the outcomes of interest. At heart, these shares are disciplined by the (complex) interaction between initial conditions and suitability through the repeated, dynamic application of the mapping f: for this reason, we condition our analysis on location amenities and the initial geography of settlement. Specification (5) *also* provides consistent estimates if the shifts are exogenous and numerous (Borusyak et al., 2022). The shifts would not be exogenous if the dynamics of recently-formed local communities in Palestine, e.g., their cohesion, were conducive of selective migration across different origins. Two considerations would provide some support to the identification assumption: (i) in our context, later migration inflows to Palestine are likely to be exogenous to local dynamics, as mostly resulting from push factors (Buggle et al., 2023); and (ii) even though Zionist organizations with an established presence in Palestine promoted and facilitated the emigration of Jewish communities across Europe, these pull factors are unlikely to be tied to the heterogeneous *local* dynamics of community cohesion within small settlements. However, we still condition our analysis on a quadratic function of the year of settlement to clean our shift-share design from secular trends in migrant characteristics.

A distinct concern is about the *interpretation* of the effect. We want the prediction $\widetilde{\vartheta}_j$

to capture diversity across the origins of settlers within settlements. However, Jews from different origins might also have different occupations, different propensity to speak Hebrew, different levels of wealth etc., and the shift-share prediction might correlate with other dimensions of settlement composition such that the effect of diversity would conflate the impact(s) of other features characterizing the successive migrant waves.

Figure 5. Predicting diversity across settlements—balance tests.



(a) Baseline (b) Additional

Notes: This Figure reports the standardized estimate of the shift-share prediction on various settlement characteristics (with 10, 5 and 1% confidence intervals). A unit of observation is a settlement. In panel (a), the dependent variables are: the standardized Herfindahl index (row 1); the standardized share of settlers who do not speak Hebrew (row 2); the standardized share of settlers from Eastern Europe (row 3); the standardized likelihood for the settlement to be a kibbutz (row 4); the standardized share of settlers who are farmers (row 5); the standardized share of white collar occupations (row 6); a standardized measure of (log) capital in 1945 (row 7); a standardized measure of market access to other Jewish settlements in 1947 (row 8). In panel (b), the dependent variables are: the standardized Herfindahl index (row 1); the standardized yield for wheat (row 2); the standardized yield for citrus (row 3); the standardized yield for vegetables (row 4); the standardized measure of market access to other Jewish settlements in 1920 (row 5); the standardized measure of market access to other Jewish settlements in 1929 (row 6); the standardized measure of market access to other Jewish settlements in 1938 (row 7); a standardized measure of Arab villages density around the settlement (row 8). The explaining variable is a standardized measure of uniformity based on (i) the timing of settlement creation as predicted by the "frontier expansion" algorithm and (ii) a Herfindahl index of yearly migrant flows. All specifications include the variables used in the "frontier expansion" algorithm (see Section 3.1): all separate geographic characteristics used to capture suitability (except if used as dependent variable); and the initial allocation of settlements and travel time to settlements around 1920. We also include a trend and quadratic term in the year of establishment, as well as population-decile fixed effects and fixed effects at the level of sub-districts as in column (2) of Table 2. Standard errors are clustered at the sub-district level.

We already test for the relevance of the instrument in Table 2. To provide support for its exogeneity and its "interpretation", we perform a balance test in Figure 5 (a), displaying the correlation between the predicted measure of uniformity and other settlement characteristics which may act as confounders: the share of settlers speaking a foreign language as first language; the share of settlers from Eastern Europe; the likelihood for the settlement to be a kibbutz; the shares of farmers and white collar occupations; a measure of capital in 1945; and a measure of market access to other Jewish settlements in 1947. Reassuringly, we find that our predicted measure of uniformity does not significantly predict the education, occupation and origin of the average settler, or the settlement type. In panel (b), we consider the correlation with agricultural yields, market access measures over time, and proximity to Arab settlements. While we do find some correlation with measures of market access, these standardized effects are much

smaller than our first-stage estimates—the standardized effect of our instrument on the Herfindahl index is about 2.2, all the other standardized effects range between 0 and 0.6 in absolute value—and we do condition on centrality to trade routes, agro-climatic conditions, and distance from Arab settlements in our baseline specification. We nonetheless provide robustness checks conditioning our specification on all other dimensions.

4 Community building within diverse settlements

This section estimates the impact of settlement diversity on community building, as measured by: (a) the local nurturing of future leaders during the Mandate; (b) community cohesion; and (c) inter-community relations with Arab neighbors.

4.1 The nurturing of societal leaders

The nurturing of societal leaders Local institutions in the Yishuv were instrumental in nation-building via the organization of socio-economic and political activities, with one notable outcome being the emergence of prominent societal leaders.

In panel A of Table 3, we investigate the relationship between the Herfindahl index of origins across all settlers (ϑ_j)—a measure of uniformity at the settlement level—and the rise of societal leaders. More specifically, our outcome of interest is constructed as the number of leaders having resided in each settlement.[23] As this measure depends on population at the settlement level, we control for population-decile fixed effects in addition to our baseline controls which are: all variables used in the "frontier expansion" algorithm, including all geographic characteristics used to capture suitability and the initial allocation of settlements and travel time to settlements around 1920; a trend and quadratic term in the year of establishment; and fixed effects at the level of sub-districts. Finally, we instrument ϑ_j with our shift-share measure based on the "frontier expansion" prediction and a Herfindahl index of yearly immigrant flows ($\widetilde{\vartheta}_j$).

We find that diverse settlements (i.e., those with a low Herfindahl index) are significantly more likely to produce prominent leaders. Specifically, a 0.10 higher Herfindahl index—roughly 40% of a standard deviation—is associated with 0.44 fewer leaders in a settlement (column 1). This negative association holds across all leader types: military (-0.16 leaders, column 2), political (-0.14, column 3), and cultural (-0.14, column 4). Given that the average settlement hosts 0.5 (winsorized) leaders, these effects are large.

In panel B of Table 3, we analyze 3,340 minutes of Knesset sessions from 1949 to 1977, comprising about 20,000 individual interventions, to classify the "type" of two-

[23] In our baseline regression, we winsorize the number of military, political, and cultural leaders in each settlement to the 99%-tile (a threshold applied to each type separately). We provide a sensitivity analysis in Appendix B.2, together with additional normalization by the settlement population.

Table 3. Diversity and the rise of societal leaders.

Leaders	All (1)	Military (2)	Political (3)	Cultural (4)
Panel A: All leaders				
Herfindahl index	-4.377	-1.589	-1.395	-1.393
	(0.798)	(0.366)	(0.429)	(0.450)
Observations	316	316	316	316
F-stat	55.31	55.31	55.31	55.31
Mean dep. var.	0.485	0.200	0.227	0.058
Leaders	Tone (+) (1)	Tone (-) (2)	Cohesion (+) (3)	Cohesion (-) (4)
Panel B: Leader type (using Knesset interventions)				
Herfindahl index	-1.666	-1.029	-1.633	-1.062
	(0.475)	(0.282)	(0.482)	(0.370)
Observations	316	316	316	316
F-stat	55.31	55.31	55.31	55.31
Mean dep. var.	0.088	0.168	0.082	0.174

Notes: A unit of observation is a settlement. In panel A, the dependent variable is the number of leaders having resided in the settlement: military, political, and cultural leaders in column (1); military leaders in column (2); political leaders in column (3); and cultural leaders in column (4). In panel B, the dependent variable is the number of leaders having resided in the settlement: in the top quartile of leaders for constructive and optimistic rhetoric at the Knesset (1949–1977, column 1), and in the bottom 75% (column 2); in the top quartile for referencing community cohesion (column 3), and in the bottom 75% (column 4). Details on data collection, processing, and description of these interventions are provided in Appendix A.4; additional outcomes (other issues and mentions of Arab-Jewish relations) are reported in Appendix B.4. The explaining variable is a Herfindahl index of origins across all settlers; the instrument is a measure of uniformity based on (i) the timing of settlement creation as predicted by the "frontier expansion" algorithm and (ii) a Herfindahl index of migrant flows. All specifications include the variables used in the "frontier expansion" algorithm (see Section 3.1): all separate geographic characteristics used to capture suitability; and the initial allocation of settlements and travel time to settlements around 1920. We also include a trend and quadratic term in the year of establishment, as well as population-decile fixed effects and fixed effects at the level of sub-districts. Standard errors are clustered at the sub-district level and reported between parentheses.

thirds of the political leaders in our sample—representing roughly 32% of all Knesset interventions during that period. Since our empirical strategy does not accommodate well sub-sample analyses restricted to settlements with at least one leader, we consider our baseline specification with leaders of different "types" as dependent variables—thus mixing an intensive margin (the number of *speakers*) with the nature of their intervention (their conditional *tonality*). In columns (1) and (2), we distinguish between "Tone (+)" leaders—those in the top quartile for constructive and optimistic rhetoric—and "Tone (-)" leaders—those in the bottom 75%. Less diverse settlements produce fewer speakers in general, and thus fewer speakers of both types, but much fewer constructive speakers: a 0.10 increase in the Herfindahl index is associated with 0.16 fewer positive-toned leaders (175% of their average incidence) and 0.10 fewer negative-toned leaders (60% of their average incidence). In columns (3) and (4), we classify leaders by their emphasis on com-

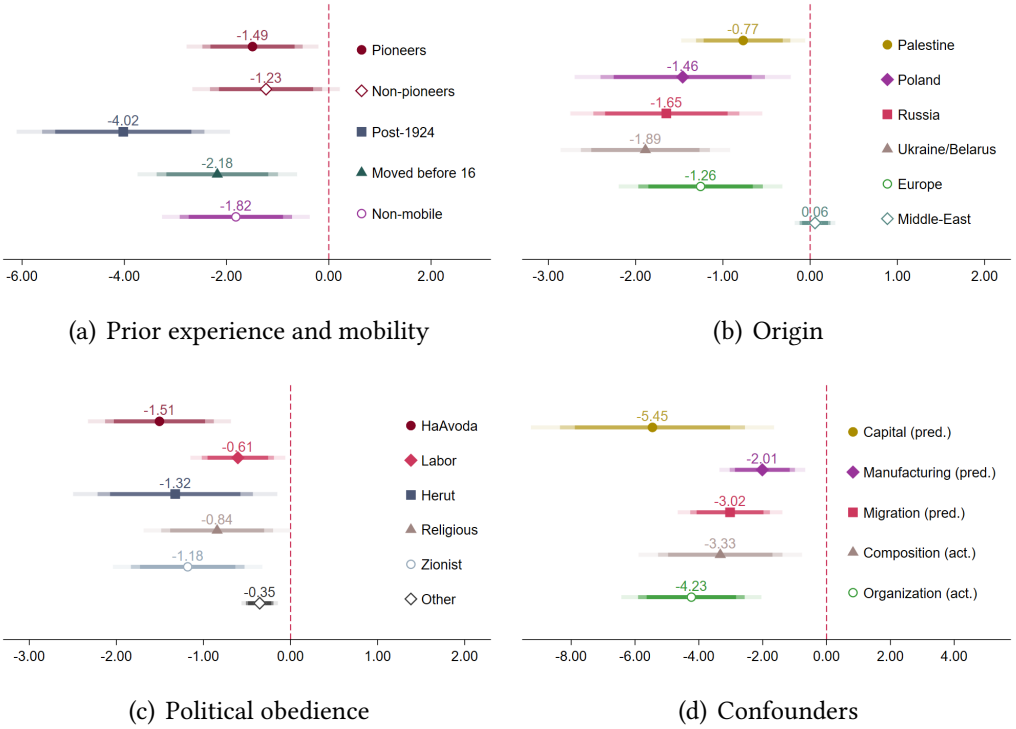
munity cohesion: "Cohesion (+)" leaders, in the top quartile for referencing community cohesion, and "Cohesion (-)" leaders, in the bottom 75%. Again, less diverse settlements are significantly less likely to produce leaders who later emphasize community cohesion. A detailed description of these interventions is provided in Appendix A.4, with a more in-depth analysis of topics, tone, and mentions of Arab-Jewish relations in Appendix B.4 (see, e.g., Figure A12).

The selection of leaders across settlements Our shift-share design mostly accounts for the endogeneity of aggregate settlement composition. However, it does not directly identify *why* diverse communities are more likely to host societal leaders. At least three plausible mechanisms could drive this relationship. [C1] Diverse settlements might attract differently skilled individuals. For example, the most able or ideologically committed individuals might be dispatched to, or later sort into, these communities. [C2] Individuals might gain influence from being associated with diverse settlements. [C3] Diverse communities might be more likely to produce leadership qualities among similarly selected settlers (e.g., because demand for leadership is higher or because it provides first-hand experience). While we cannot exclude that all these explanations do play a role in our estimates, possibly in conjunction with each other, our preferred interpretation is the third. In Section 5.2, we show that diverse communities do present unique challenges to their inhabitants, potentially enabling the emergence of skilled politicians and opinion leaders [favoring C3 over C2]. We further support this interpretation here through a systematic analysis of leader selection across settlements [C1].

Our societal leaders differ widely in ideology, preparedness upon arrival, and upbringing, raising the question of whether their allocation across settlements reflects these characteristics. In a first exercise, we investigate whether leaders with different levels of prior experience or ideological commitment systematically selected into specific types of settlements, e.g., based on their composition or cohesiveness. This issue is examined comprehensively in Appendix B.4, alongside robustness checks related to our leader selection criteria and methods of assigning them to settlements when multiple residences were recorded. Summary results are shown in panel (a) of Figure 6. We find that our diversity effect is robust to: (i) excluding the 64 selected leaders with prior experience as pioneers or political activists and only keeping *non-pioneers*; (ii) excluding the 70 early arrivals, who were more likely to have received formal training, and keeping *post-1924* migrants; (iii) only considering leaders having settled in their assigned communities *before the age of 16*; and (iv) excluding "mobile" leaders, thereby reducing concerns about the importance of post-arrival mobility and selective leader assignment.[24]

[24]Appendix B.4 provides a more comprehensive analysis, notably using a "probabilistic" assignment

Figure 6. Diversity and the rise of societal leaders—a selection of complementary evidence.



(a) Prior experience and mobility

(b) Origin

(c) Political obedience

(d) Confounders

Notes: This Figure provides a selection of robustness checks and additional evidence, which are more comprehensively covered in Appendix B.2 and Appendix B.4. Panel (a) displays the equivalent to column (1) of Table 3 in a specification similar to Equation (5), but where the dependent variable is replaced with: the number of pioneers; non-pioneers; leaders having arrived after 1924; leaders having settled in their assigned communities before the age of 16; and leaders with only one recorded location across the whole period. Panel (b) replaces the dependent variable with the number of leaders from different regions of origins (Palestine, Poland, Russia, Ukraine/Belarus, the rest of Europe, Middle-East). Panel (c) replaces the dependent variable with the the number of leaders affiliated with HaAvoda, other labor parties, Herut (conservative), religious parties, Zionist parties and other parties. Panel (d) displays the estimates for all leaders (equivalent to column 1 of Table 3) in a specification similar to Equation (5), but where we add the following controls: the predicted share of capitalists; the predicted share of settlers with prior occupation in manufacturing; the predicted migration flows; the full, actual occupational structure, the share of males who do not speak Hebrew, and the share of illiterate males; and settlement types (Kibbutz, Moshav, and Other).

In a second exercise, we analyze detailed leader biographies, focusing on their origins and later ideological affiliations. Specifically, we examine how the diversity effect varies by region of origin, by political party affiliation—a key ideological marker, particularly in relation to attitudes toward the Arab population, and by post-1948 governmental involvement. We report a selection of these findings in Figure 6 and leave the more detailed analysis to Appendix B.4. We find that the diversity effect is broadly consistent across

of leaders across their different residences, better accounting for the length of stay across residences, or using an externally-curated sample of 6,000 influential settlers with shorter biographies. In Appendix A.1, we also discuss (i) how we identify meeting participants across settlements, (ii) provide a list of the most frequent contributors (likely to be prominent figures within the Yishuv), and (iii) show that the overlap between these contributors and our identified leaders is limited to about 15 individuals (e.g., David Ben-Gurion, Israel Bar-Yehuda, Gideon Ben-Israel, David Elazar, Yehoshua Hankin, Pinhas Lavon, and Yitzhak Tabenkin). The vast majority of leaders in our analysis were *not* prominent figures during the Yishuv period.

regions of origin (panel b of Figure 6) and major political groupings of the time (panel c of Figure 6), including HaAvoda/Mapai led by Ben-Gurion, other labor parties, liberals, and conservatives. Exceptions include the small number of leaders from the Middle East.[25]

These exercises reduce concerns that the relationship is driven by a selection of migrants with stronger leadership abilities into diverse areas or the different ideologies of migration waves [C1]—rather than the diversity of these locations fostering the development of leaders [C3].

Identification and interpretation of the effect Our identification strategy relies on the assumption that the predicted uniformity measure is orthogonal to local settlement dynamics (*exogeneity*)—a condition which is satisfied if migrant inflows to Palestine (the "shifts") are exogenous. However, the "shares" of our shift-share design partly rely on geography interacted with the initial settlements at the start of the British Mandate, and one might worry that our "frontier expansion" algorithm picks up other, subtle geographic differences that are not well accounted for in the baseline regression (controlling for all separate geographic factors and for functions of the year of settlement creation). For instance, although more diverse settlements are not better located *ex-ante*, they could end up better located *ex-post*, given the evolving spatial distribution of the Jewish population during the Mandate. We show the (lack of) importance of these controls in Appendix B.2 (Figure B4), where we successively condition on geography, initial settlements, sub-districts fixed-effects, migrant shares by origin in 1945, settlement type (kibbutz), market access (in 1947) and overall immigrant flows (allowing us to disentangle the diversity effect from a scale effect generated by immigration levels).

Another challenge concerns the *interpretation* of the diversity effect. Our instrument is indeed based on the heterogeneity of immigrants' countries of provenance over time. Settlements with an exogenously diverse population could however differ from other communities of Mandatory Palestine along other characteristics than just birthplace diversity. For instance, diverse migration waves might generate different, observable settlers' characteristics: overall prevalence of certain regions of origin, strength of the majority group, skills, former occupations, or language—traits which may correlate with leadership ability.[26] To address this interpretation issue, we condition our estimates on:

[25] In Appendix B.4, we further find that the effect remains large for political leaders with high responsibilities (Table B8); we consider standardized leader "scores" based on the number of governmental roles and spells at the Knesset, the length of their biographies, and the number of publications in Table B9; and we investigate whether the effect derives from leaders coming from dominant groups within settlements or from a leadership competition between a majority group and the second-most-prominent group, and whether diverse settlements lead to the rise of religious leaders.

[26] Another, less observable confounder would be the ideology and training of those settlers, which could have contributed to or hindered the rise of leaders and the creation of a shared identity within the settlement. Ideology manifested itself in several ways: (i) leadership experience prior to migration, in

predicted settler wealth and occupation from a shift-share design similar to Equation (4); predicted migration incidence; actual composition of each settlement in terms of occupation, language, and cultural background (e.g., prevalence of settlers from the Pale of settlement—characterized by a "shtetl culture" and different socio-cultural background); and organizational structure (e.g., kibbutz type and affiliation). We report selected tests in panel (d) of Figure 6, and provide a comprehensive analysis in Appendix B.3. We find that our leadership effect is not mitigated by differences in prosperity, skill levels, agricultural composition, or settlement organization.[27]

Sensitivity analysis We conclude with an extensive set of robustness checks and sensitivity analyses (detailed in Appendix B.2), further validating our baseline specification: (i) we show the robustness of our first-stage prediction to conditioning on settlement composition (occupations, education, separate origins), capital, settlement types and measures of market access in Appendix B.1; (ii) we report OLS estimates in Table B3; (iii) we consider alternative instruments, i.e., a polarization index, a shift-share prediction based on larger regional groupings (reflecting the fact that Jewish identities did not necessarily coincide with country borders, e.g., within the Pale of Settlement), and a Herfindahl index of predicted settlement composition instead of a weighted average of Herfindahl indices (see Tables B1 and B4 for the first- and second-stage specifications); (iv) we consider alternative standardizations of leadership counts (e.g., showing that our results are not driven by a few settlements with many leaders, panel a of Figure B4); and (v) we show in Figure B4 that our results are robust to using different approaches to inference (e.g., clustering at a less granular level or using Conley standard errors to account for spatial auto-correlation) and to smaller/larger sets of geographic controls.

4.2 Community relationships and conflicts

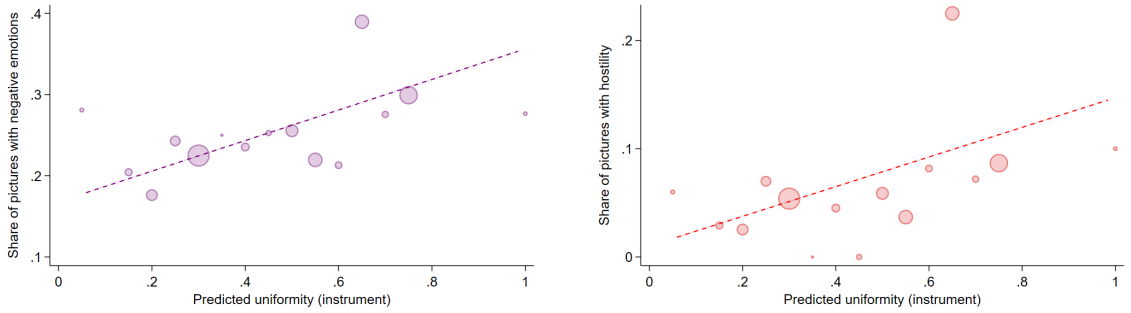
Other visible manifestations of functioning local institutions are community cohesion and more peaceful relationships with nearby Arab villages. We capture these relationships through imperfect proxies prior to the war and during the war.

Community cohesion and inter-ethnic relationships before the war We exploit two unconventional data sources to assess community cohesion and inter-ethnic rela-

pioneering movements or in political parties; (ii) the choice of settlement types upon migration (e.g., the moshav, kibbutz, moshava or an urban destination); (iii) the motivation behind migration, e.g., ideology rather than persecution; and (iv) different commitment with respect to the "Negation of the Diaspora", which involved a willingness to abandon pre-aliyah identities (Engel, 2021).

[27] We will however show in Section 5 that settlement structure does matter in explaining treatment *heterogeneity*, i.e., the size of the diversity effect, in part because the kibbutz generally appears to be a distinct contributor to community cohesion and to the nurturing of societal leaders.

Figure 7. Photographs and community relationships.



(a) Emotions (b) Hostility

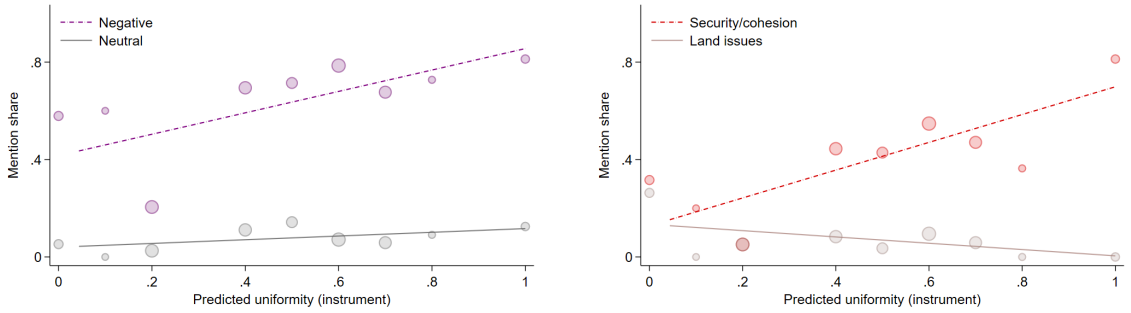
Notes: Panel (a) displays the correlation between our shift-share uniformity prediction, $\widetilde{\vartheta}$, and the average negative emotion score (summing the probability scores for anger, disgust, fear, and sadness). The data is based on 3,385 pictures which can be assigned to one of about 45 settlements of our sample. For each picture, the emotions are a probabilistic vector as provided by the large language model (anger, amusement, awe, contentment, disgust, excitement, fear, sadness—see Appendix A.3 for examples). Panel (b) displays the correlation between our shift-share uniformity prediction, $\widetilde{\vartheta}$, and the average probability that interactions are classified as hostile within the 1,500 pictures depicting meaningful social interactions. For exposition purposes, we group observations in dots by deciles of the shift-share uniformity prediction, $\widetilde{\vartheta}$, weighted by the number of mentions, and we show the best linear fit(s).

tionships, motivated by data scarcity and the lack of statistical coverage of Arab villages.

First, we analyze a set of 3,400 photographs that can be linked to 45 communities in our sample during the relevant period (excluding Tel Aviv and Jerusalem; see Appendix A.3). These images capture a variety of scenes, including daily life (e.g., harvests), minor events (e.g., the opening of a factory or training school), and major incidents (e.g., riots). Due to incomplete settlement coverage, we adopt a simplified empirical approach. In panel (a) of Figure 7, we correlate the average negative emotion score of photographs—defined as the probability that an image is classified as conveying negative affect—with our shift-share measure of settlement uniformity, $\widetilde{\vartheta}$. We find that more diverse settlements are significantly less likely to be associated with negative imagery: the average negative emotion score rises from approximately 0.18 in the most diverse settlements to 0.35 in the least diverse. This finding could be due to selection in the subject of the photograph. To address this concern, panel (b) of Figure 7 restricts the analysis to images depicting social interactions: the probability of hostile interactions increases from 0 in the most diverse settlements to 0.12 in the least diverse.

The previous social interactions most likely arise from *within* Jewish communities, but might also capture tension *between* Arab and Jewish villages. To capture inter-community relations, we conduct a contextual topic analysis of meeting minutes. As described in Section 2, these meetings took place in approximately 30 Jewish settlements between 1930 and 1947. While hosted locally, participants often came from other locations, and discussions frequently addressed issues pertaining to multiple settlements.

Figure 8. Settlement meetings and inter-community relationships.



(a) Tone (b) Subject

Notes: This Figure displays the correlation between our shift-share uniformity prediction, $\widetilde{\vartheta}$, and the share of mentions of Arab neighbors, within meeting minutes, that are associated with a negative tone (urgent, worried, critical—dashed line) versus a neutral tone (plain line) in panel (a) and the share of mentions related to the topics of security and community cohesion (dashed line) versus land issues (plain line) in panel (b). The large language model produces 270 mentions across 60-70 collated minute reports (and about 6,000 pages) referring to 40 settlements of our sample. The settlements are: Aiyanot (Ramat David), Alonim, Ashdot Ya'aqov, Beer Ya'aqov, Beit Hashitta, Beit Oren, Caesarea, Dalhalmia/Gesher, Deganya A, Deganya B, Ein Gev, Ein Harod, Ein Hayam, Geva, Ginegar, Givat 'Ada, Givat Brenner, Givat Hashlosha, Hadera, Haifa, Hulda, Kefar Hahoresh, Kefar Hammakkabi, Kefar Sava, Kinneret, Ma'alei Hahamisha, Mazkirut HaPeila, Menahemya, Mishmarot, Na'an, Ramat Yohanan, Rehovoth, Revivim, Sede Nahum, Shahariya, Shefayim, Tel Yosef. For exposition purposes, we group observations in dots by deciles of the shift-share uniformity prediction, $\widetilde{\vartheta}$, weighted by the number of mentions, and we show the best linear fit(s).

We use the latest version of a large language model (GPT-4o) to identify instances where specific settlements are explicitly referenced in relation to Arab communities. For each mention, we extract the topic (land acquisition, security, economy, agriculture, or community cohesion), tone (urgent, worried, critical, neutral, optimistic, or constructive), general positivity (from 0 to 100), and the broader context of the discussion. Given the limited number of observations, we adopt a simplified empirical approach, correlating the tone and topic of these references with our instrumental variable predicting Jewish settlement composition. Panel (a) of Figure 8 shows that the share of negative-toned mentions increases from about 40 percentage points to 80 percentage points between the most diverse and least diverse settlements. Panel (b) shows that it partly reflects the nature of these discussions: the increasing negativity with declining diversity mirrors a corresponding rise in discussions focused on security and/or cohesion issues.[28]

[28]We require the large language model to produce quotes to justify the inclusion of entries. Two quotes illustrate the disparity in the treatment of inter-ethnic relations. Ein Hayam had a proactive approach to try and form links with the neighboring Arab villages (1945), an endeavor that is still present in the local institutional setting nowadays: "The issue of neighborly relations is not new and there is no need to elaborate on it. It's been 21 years since the Agency tried nationally to regulate this issue. From the course of events, it became clear that the relations between the Arab and Jewish settlements were estranged and it was necessary to start dealing with this systematically. [...] The plan was broad and large but it was not to be reduced. 1) Educational part 2) Research part 3) Practical field. Everyone recognizes the need for each settlement to have some percentage of Arabic speakers." By contrast, discussions in Ashdot Ya'aqov around 1940 mention: "In general, our greatest danger, the most important agent of Nazism, of the extermination of the Jews, of the danger to Zionism will be our Arab neighbors. Not necessarily in the

The previous evidence offers valuable insights using novel data sources. It does not, however, accommodate our baseline empirical specification. We now turn to more systematic evidence on Arab/Jewish relations, focusing specifically on the wartime period.

Conflicts and inter-ethnic relationships during the war To quantify inter-ethnic relationships during the war, we rely on 100 informal agreements between Jewish and Arab settlements (Azoulay, 2014) and data on depopulated villages, accounting for civilian casualties among the Arab population. Indeed, while military plans were centralized, a non-negligible share of soldiers engaged in military operations came from local settlements and brigades had some degree of discretionary power over the intensity of the attack and whether these attacks should target civilians or not.

Table 4. Conflicts, inter-ethnic relationships, and the composition of early settlements.

	Civilian casualties	Attack	Depopulated
War events	(1)	(2)	(3)
Panel A: War in nearby Arab villages (within 10 kms)			
Herfindahl index	1.437	0.130	0.129
	(0.453)	(0.185)	(0.165)
Observations	316	316	316
F-stat	70.49	70.49	70.49
	Peace	Cooperation	Reconciliation
Pacts	(1)	(2)	(3)
Panel B: Pacts between Jewish settlements and nearby Arab villages			
Herfindahl index	-0.571	-0.216	-0.427
	(0.300)	(0.336)	(0.139)
Observations	316	316	316
F-stat	55.31	55.31	55.31

Notes: A unit of observation is a settlement. In panel A, the dependent variable is the probability to be exposed to the following events involving at least one Arab village within 10 kms: having civilian casualties in column (1); being attacked in column (2); and being depopulated in column (3). In panel B, the dependent variable is the number of pacts mentioning peace (e.g., a peace agreement or a non-aggression pact), cooperation and prevention of a conflict (e.g., declaration of solidarity or warning about the arrival of troops), or reconciliation (e.g., apologies) between each Jewish settlement of our sample and neighboring Arab villages. The explaining variable is a Herfindahl index of origins across all settlers; the instrument is the shift-share uniformity measure, $\widetilde{\vartheta}$, based on the timing of settlement creation as predicted by (i) the "frontier expansion" algorithm and (ii) a Herfindahl index of yearly migrant flows. All specifications include the same controls as in Table 3, with the addition of the local Arab population and the presence of military Arab troops in nearby villages in panel A. Standard errors are clustered at the sub-district level and reported between parentheses.

In panel A of Table 4, we investigate the relationship between our Herfindahl index of origins within Jewish settlements and the intensity of wartime violence affecting nearby Arab communities. Specifically, we construct the probabilities that at least one

city. If there is a serious military struggle here - then the greatest danger will be through every village."

Arab settlement within a 10-kilometer radius was (i) the site of civilian casualties, (ii) exposed to a military attack, or (iii) depopulated. Our baseline specification follows that of Table 3, with two additional controls: local Arab population density and the presence of Arab military forces. We find a gradient linking settlement homogeneity to civilian casualties: A 0.10 higher Herfindahl index within a Jewish settlement is associated with a 0.14 higher probability of proximity to Arab civilian casualties (column 1, panel A).[29] The diversity effect is much lower and not statistically significant for (ii) military attacks and (iii) depopulated Arab villages, reflecting that the choice of targeted villages was mainly driven by top-down military and strategic considerations (consistent with directives outlined in Plan Dalet) rather than the nature of inter-ethnic relationships, and that depopulation operations were so widespread that almost all our settlements were in the proximity of such events.

In panel B, we analyze the incidence of cooperative pacts between Jewish settlements and neighboring Arab villages. These include peace agreements, declarations of solidarity and warnings about incoming attacks, and reconciliatory gestures. Approximately 50 settlements in our sample of 316 engaged in such actions; and we treat all pacts symmetrically, regardless of which side initiated them. We show that a 0.10 higher Herfindahl index within a settlement is associated with 0.055 fewer peace agreements with Arab neighbors (column 1) and 0.042 fewer apologies to/from the other party (column 3). The effect on cooperation/prevention is more muted. In general, however, and summing these estimates as we do in Appendix B.5, a 0.10 higher Herfindahl index within a settlement is associated with 0.11 fewer pacts between communities.

Interpreting reduced violence in our context of nation-building is not straightforward. First, does reduced violence necessarily advance nation-building? Jewish communities and our societal leaders held diverse views on coexistence with Arabs. A strand of Zionist thought, exemplified by Haim Arlosoroff, emphasized integration and cooperation, framing peaceful coexistence as essential to Zionism and Palestine's development. Although initially influential, this approach was gradually displaced by more aggressive positions on territorial expansion during the Mandate and by segregationist policies in the early Israeli state (implemented by political leaders in our sample, see Kafkafi, 1998). Thus, the ex-post dynamics of the region suggest that settling disputes peacefully might not have been instrumental in the eventual development of Israel. Second, does diversity directly promote pacifism within communities, or does it operate indirectly by fos-

[29] In Appendix B.5, we look at treatment heterogeneity depending on whether the attack leading to civilian casualties was executed by local or national forces, and we find that the results are exclusively driven by locations where Plan Dalet was carried out by local forces. We also consider dummy variables for being exposed to non-combatant casualties within a radius of 5, 15 or 20 kilometers around each Jewish settlement, and we find a spatial decay in this effect.

tering more constructive local leadership? There is support for both channels. In line with the contact hypothesis, cooperative interactions within diverse Jewish communities may have shaped more tolerant attitudes toward out-groups, which then extended to neighboring Arab villages. Similar mechanisms have been documented in other settings, e.g., with the higher tolerance of Muslims after performing the Hajj (Clingingsmith et al., 2009), or in controlled laboratory experiments (Pettigrew and Tropp, 2006). Alternatively, internal dynamics of diverse settlements may have nurtured leaders with more constructive orientations, who in turn eased relations with the Arab population.[30] Indeed, we find support for this latter channel: leaders from more diverse communities were more likely to adopt conciliatory positions, as reflected in Knesset debates on Arab-Jewish relations between 1949 and 1977 (see, e.g., Figure A12 in Appendix B.4).

5 The mechanics of community building

The previous section finds an important role for the *making* of local communities, with diverse settlements more likely to nurture societal leaders and foster better relationships with Arab neighbors. In this section, we shed some light on the *functioning* of these local communities to better understand the mechanics of community building.

5.1 Community building within diverse, differently-organized settlements

We proceed in two steps. In this first step, we look at the heterogeneous social organizations of Jewish communities, shaping the nature of social interactions, community cohesion, and the formation of a collective identity. These rural settlements markedly differ in their structure and institutions: the kibbutz, an emblematic institution of Mandatory Palestine, had a democratic structure guiding decision-making processes and was affiliated to ideological movements promoting egalitarian principles and communal lifestyle; the moshav was a less integrated, cooperative agricultural community where equal-size farms would be privately-owned and education would be less communal (Weintraub et al., 1969); and the moshava was an agricultural settlement with private property over land and its proceeds, usually privileged by earlier migration waves to Ottoman Palestine. The structuring of rural communities into these different models is likely driven by endogenous dynamics and the involvement of larger ideological movements; our focus will thus be on the effect of diversity *within a given structure*.[31] In a second step, we will

[30] Conversely, security concerns may have heightened the need for internal cohesion in settlements established in strategic locations (Diao, 2021).

[31] We discuss more explicitly the differences in settlement characteristics—e.g., geography, connectedness, composition—across these different modes of organization in Appendix B.6. In particular, we show that kibbutzim appear to be negatively selected with respect to their ex-post geography in Mandatory Palestine, that they are (far) less likely to host white-collar workers, but that: (i) they tend to be similarly

further focus on the nature of community building in the kibbutz.

We systematically classify our 316 settlements depending on their organization and main ideologies. We then estimate treatment heterogeneity within a simple transformation of our baseline specification (5), distinguishing between the kibbutz (k, 134 communities), and the other settlements (o, the moshav, cities or urban areas, and other agricultural settlements). Letting j denote a settlement of type $m \in \{k, o\}$, we estimate the heterogeneous effect of local diversity on nation-building, y_j, as follows:

$$y_j = \alpha + \beta_o \vartheta_j + \beta_k \vartheta_j \times \mathbb{1}_{m=k} + \gamma \mathbf{X}_j + \delta \mathbf{M}_j + \varepsilon_j, \tag{6}$$

where $\mathbb{1}_{m=k}$ equals one if the settlement is a kibbutz. We instrument ϑ_j and its transformation, $\vartheta_j \times \mathbb{1}_{m=k}$, with the shift-share prediction, $\widetilde{\vartheta}_j$, and its transformation, $\widetilde{\vartheta}_j \times \mathbb{1}_{m=k}$. In order to account for heterogeneity across settlement structure, we further control for $\mathbf{M}_j$, a set of organization fixed effects—distinguishing between: different ideological kibbutz movements (Kibbutz Artzi; HaKibbutz HaMeuhad; and religious kibbutzim); the Moshavim Movement; other Moshav; and unaffiliated settlements.

Table 5. Diversity, the rise of societal leaders and inter-ethnic relationships—treatment heterogeneity.

	All leaders	Mil. lead.	Pol. lead.	Cul. lead.	Casualties	Pacts
	(1)	(2)	(3)	(4)	(5)	(6)
Herfindahl	-2.235	-0.854	-0.367	-1.015	1.283	-0.385
	(1.349)	(0.478)	(0.592)	(0.434)	(0.631)	(0.553)
Herfindahl × kibbutz	-9.185	-2.086	-5.387	-1.712	1.892	-2.763
	(5.438)	(2.333)	(2.408)	(1.235)	(2.487)	(1.213)
Observations	316	316	316	316	316	316
F-stat (Herfindahl)	78.08	78.08	78.08	78.08	78.08	78.08
F-stat (Interaction)	8.98	8.98	8.98	8.98	8.98	8.98

Notes: A unit of observation is a settlement. The dependent variable is: the number of leaders having resided in the settlement in column (1); military leaders in column (2); political leaders in column (3); cultural leaders in column (4); the probability of civilian casualties in nearby Arab villages within 10 kms in column (5); and the total number of pacts between each Jewish settlement of our sample and neighboring Arab villages in column (6). All specifications include the same controls as in Table 3, with the addition of the local Arab population and the presence of military Arab troops in nearby villages. Finally, all specifications include dummies for the different settlement organizations: Kibbutz Artzi, HaKibbutz HaMeuhad, the Moshav Movement, other moshavim, Chever Hakvutzot, other religious kibbutzim, and other structures including major cities. F-statistics are derived using the weak instrument F-test proposed in Sanderson and Windmeijer (2016). Standard errors, in parentheses, are clustered at the sub-district level.

Table 5 reports the estimates (β_o, β_k) for the following outcomes: the number of societal leaders in columns (1)-(4); the probability to record civilian casualties in neighboring Arab settlements during the war; and the total number of pacts with neighboring Arab villages. Table 5 shows that the diversity effect observed for societal leaders and inter-

diverse as other communities at any given date of settlement creation; and (ii) conditioning on the previous differences in the *making* of local communities does not affect the estimates from specification (6).

community relationships is mostly confined to the kibbutz as settlement structure: a 0.10 higher Herfindahl index is associated with 0.22 fewer leaders within non-kibbutz settlements, but 1.13 fewer leaders in kibbutzim, with the most striking heterogeneity being observed for the rise of *political* leaders and the formation of pacts with Arab villages.[32]

In summary, diverse communities appear conducive to the emergence of societal leaders and improved relations with Arab neighbors, but primarily within a specific institutional framework that features integrated internal organization.[33] At heart, the implications of diversity for community dynamics remain ambiguous: interactions within fractionalized communities of many small groups can facilitate the creation of a shared identity (Bazzi et al., 2019) and coordination mechanisms often emerge in response to frictions arising from cultural, linguistic, or ideological differences (Alesina and La Ferrara, 2005; Habyarimana et al., 2007; Greif, 1993), but diversity may impose coordination costs, particularly in settings with decentralized interactions.[34] The kibbutz—with its highly integrated structure and the active engagement of its grassroots community—may have offered the ideal institutional setting to reduce coordination frictions and institutionalize cooperation. The next section studies how such communities identified, confronted, and resolved internal challenges, and how these processes may have facilitated coordination and the development of societal leadership.

5.2 Community cohesion and the grassroots of nation-building in the kibbutz

We use minutes from settlement meetings and secondary sources to discuss local challenges, community cohesion, and the grassroots of nation-building within the kibbutz.

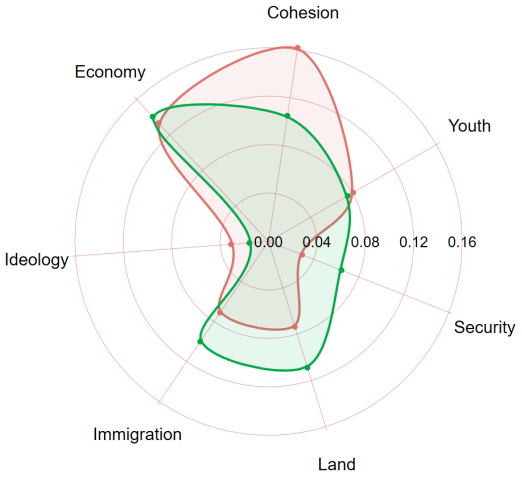
[32]Note, for the sake of transparency, that there is a direct, positive kibbutz effect on the production of leaders, irrespective of settlement composition and their umbrella organization. We further investigate the rise of *different* leaders across settlement structure in Appendix B.7, where we look at the heterogeneous production of leaders along their governmental roles, their political affiliation, and their birthplace. The diverse kibbutz is shown to produce influential leaders of the most frequent political obedience of the times and from the main immigrant origins, except for religious parties, conservative organizations, and individuals from the Middle-East. These exceptions are not surprising: Mizrahi Jews, or more conservative settlers, were more likely to opt out of the kibbutz, its extreme egalitarian principles, and its communal education and rather choose to settle in the moshav or the moshava.

[33]While kibbutzim are often portrayed as sites of egalitarian cooperation, recent scholarship highlights how narratives of coexistence were entangled with processes of Palestinian dispossession. In particular, Sabbagh-Khoury (2023) shows that left-wing kibbutzim of the 1930s simultaneously promoted ideals of binationalism while pursuing land practices that contributed to the Nakba.

[34]Reducing coordination costs can be achieved through various channels, e.g., by limiting residential segregation, attending the same schools (Bandiera et al., 2019), sharing the same religion (Clingingsmith et al., 2009), speaking a common language (Bazzi et al., 2019), or marrying across communities (Duncan and Trejo, 2011). In the context of our Jewish communities, however, there was limited variation across these dimensions: Hebrew was spoken in most households; Jewish children were not segregated in schools; and the population mostly shared the same religion, albeit with different strength of religious belief.

Challenges and community cohesion as inferred from meetings We exploit our textual analysis described in Section 2 to shed light on the challenges faced by communities of different making.

Figure 9. Diversity and challenges as inferred from meetings.



Notes: This Figure displays the share of entries mentioned with a negative tone (i.e., critical, urgent, worried, or neutral, but excluding: constructive, optimistic) for the following grouped topics: cohesion (community cohesion, language); immigration (absorption of immigration); ideology (ideology, and issues related to the secretariat); economy (economic questions, financial issues, and infrastructures); land (including housing, agriculture, and water); security (including questions regarding the army); and youth (including education and recruitment). The red (green) curve shows the average share of entries for settlements with above-median (below-median) predicted measure of uniformity.

To identify pressing challenges faced by communities, Figure 9 displays the normalized incidence of entries mentioned with a negative tone across the following themes: cohesion (community cohesion, language); absorption of immigration; ideology (ideology, and issues related to the secretariat); economy (economic questions, financial issues, and infrastructures); land (including housing, agriculture, and water); security (including questions regarding the army); and youth (including education and recruitment). We create the normalized incidence of these themes in settlements with above-median (red) and below-median (green) predicted measures of uniformity. Figure 9 shows that: the main challenge faced by the average community is community cohesion, followed by economic hurdles, land, and the absorption of immigrants; the diverse settlement is far less likely to worry about community cohesion and more likely to worry about other challenges, including the absorption of immigration and land issues.[35]

[35]Jewish immigrants brought diverse cultural traditions, making assimilation a key challenge. Archival documents from Yavne, Kfar Ya'bets, Tirat Tzvi, Be'erot Yitzhak, Shluchot, Beit Alfa, Afikim, Sde Eliyahu, Neveh Itan, and Beth Yehoshua explicitly mention how to integrate new migrants with existing kibbutz members; they discuss the difficulties in absorbing immigrants from different countries and debate

Table 6. Community cohesion within the kibbutz—a topic analysis.

Cohesion	All tone (1)	Negative tone (2)	Neutral tone (3)	Positive tone (4)
Herfindahl	0.135	0.210	-0.060	-0.015
	(0.085)	(0.111)	(0.038)	(0.074)
Observations	316	316	316	316
F-stat	55.31	55.31	55.31	55.31

Notes: A unit of observation is a settlement. The dependent variable is the normalized number of entries classified as mentioning "community cohesion" with: any tone (column 1), a negative tone (column 2, urgent, worried, critical), a neutral tone (column 3), a positive tone (column 4, constructive, optimistic). The explaining variable is a Herfindahl index of origins across all settlers; the instrument is a shift-share prediction based on (i) the timing of settlement creation as predicted by the "frontier expansion" algorithm and (ii) a Herfindahl index of yearly migrant flows. All specifications include the same controls as in Table 3. Standard errors, in parentheses, are clustered at the sub-district level.

In Table 6, we further focus on the variation in discussions of social cohesion across settlements. More specifically, we run our baseline specification (5) on the normalized incidence of the topic "community cohesion" within the minutes of meetings associated with a settlement. In column (1), we consider all mentions irrespective of their tone. In column (2), we focus on negative mentions (with the tones: urgent, worried, critical). In column (3), we consider neutral mentions. In column (4), we consider positive mentions (constructive, optimistic). Table 6 shows that diverse settlements are far less likely to engage with the topic of community cohesion using a negative tone: A 0.10 higher Herfindahl index is associated with a 0.0075 lower share of minutes discussing cohesion in positive/neutral fashion and a 0.021 share of minutes discussing it negatively.

These findings might seem puzzling: how can diverse communities be facing assimilation issues and yet appear more harmonious and balanced in their coverage of important challenges? We turn to these questions next.

Community cohesion, coordination and leadership within the diverse kibbutz

The grassroots role of the kibbutz in nation-building is well documented (Near, 1992, 2011). To foster cohesion and coordination, kibbutzim engaged in identity-cementing cultural practices such as folk songs, dance, songwriting, theater, ideological or political writing, most of which appeared in local journals (Near, 1992). These activities were often conducted in Hebrew, and archival records show organized efforts to teach the language to new immigrants—e.g., in Kfar Ya'bets, Afikim and Tirat Tzvi, as reported in the Kibbutz Secretariat News, 1944, 58 (41); 1945, 30 (27); 1945, 29 (32). Kibbutzim developed specific training programs, classes on Zionism, and ideological educational

whether different groups could be combined or were culturally too different—see also the Kibbutz Secretariat News, 1945, 18 (21), 1945, 21 (24), 1946, 30 (33) and 1947, 43 (46).

methods to attract migrants and create a new Jewish identity (Israel Archive, Kibbutz Secretariat News, 1947, 42 (45)). In addition to cultural practices, institutional design facilitated coordination. The kibbutz—as a successor of the shtetl—promoted mutual aid and solidarity (Spiro, 1956; Rabin, 2013; Rayman, 1981), but crucially operated under a democratic structure: general assemblies and councils enabled broad participation in decision-making. These institutions functioned as laboratories in the promotion of the "in-gathering of the exiles", in their assimilation, and in the nurturing of societal leaders.[36] However, why did such cohesion and leadership emerge in *diverse* kibbutzim?

Diversity may have amplified the demand for coordination, which the kibbutz structure was uniquely equipped to meet. First, substantial efforts were needed and provided to ensure that migrants would rapidly shed their different cultures and preferences (Weingrod, 1985), possibly making it easier for a new model of local communities to emerge (an argument of unity in diversity, see Bazzi et al., 2019). While anecdotal, the community adopting the most constructive tone during meetings is the diverse Giv'at Brenner, funded by the pacifist Enzo Sereni.[37] Second, these communities would have general assemblies, composed of representatives of its members and discussing a wide array of topics ranging from day-to-day activities to longer-run questions; the diverse kibbutz might have mechanically involved more representatives of the different groups. For example, meeting attendance was significantly higher (20–30 participants) in the diverse Ein Harod, Ginegar, Giv'at Brenner, Giv'at Hashlosha, Hulda, or Ramat David compared to the less diverse Shefayim or Deganya A. Meeting records also show a broader range of topics discussed in the former group. This widespread involvement supported the collective provision of education, health, or the organization of work and redistribution. Such inclusive participation may have nurtured leadership through two channels: (i) a *supply effect* with more individuals involved in governance, and (ii) a *demand effect* with diversity itself generating more complex challenges to resolve.

Through the local, grassroots nurturing of future societal leaders, the (diverse) kibbutz became an important factor contributing to nation-building in a heterogeneous society. To our knowledge, this represents a novel channel through which inclusive, local institutions—acting as social laboratories—can foster a culture of social cohesion and engagement and shape the wider organization of the state (Sellers et al., 2020).

[36] Political activity and leadership were encouraged in the kibbutz, with some kibbutzim giving practical training to local leaders (e.g., Ein Harod or Kfar Ya'bets).

[37] For instance, Giv'at Brenner is mentioned in 1935 as a place where: "Despite the differences of opinion we have, [...] [A]s we meet in the seminar, we act as members in a joint venture. [...] we must create a framework for mutual understanding and mutual growth; this is the most ingrained problem". In 1939, "It is worthwhile to invest efforts in the group because there are definitely good members among them who will be a blessing to the kibbutz." Together with Givat Hashlosha, Giv'at Brenner organized "Study Months", "a valuable tool in the kibbutz's cultural activities and a unifying factor for the entire society".

6 Conclusion

This paper investigates the formation of societal cohesion and the emergence of leadership in times of mass migration. Our setting is unique: immigrants, often displaced from their places of origin, settled in agricultural communities along the Palestinian frontier—within a territory lacking formal institutions and marked by tension between Jewish settlers and the local Arab population. Our evidence combines a shift-share design allocating migrant waves to the settlement frontier with novel historical sources from a data-scarce environment (e.g., meeting minutes, photographs, biographies). We find that two factors interacted to foster social cohesion and the rise of societal leaders: (i) origin-based diversity among Jewish settlers, which generated local challenges mirroring those later faced at the national level; and (ii) an emblematic institutional setting—the kibbutz—whose democratic and integrated structure replicated key functions of government. Beyond internal cohesion, we also show that diversity helped mitigate tensions with neighboring Arab communities, even as the broader institutional project remained structured around separation (Shafir, 1996).

Our analysis faces two limitations. First, although the Jewish population of Mandatory Palestine was remarkably diverse, many settlers were willing to forgo their past identities and adhere to the practicalities and ideologies of their new environment. Second, the kibbutz was a singular institution, combining communal living with participatory governance—two features rarely observed together in migrant communities. While the Jewish experience in Mandatory Palestine had distinctive features, it was part of a broader pattern in which migrant and displaced groups sought to build new communities from below. A striking parallel is nineteenth-century Liberia, where Americo-Liberian settlers—descendants of enslaved Africans with diverse origins (Igbo, Yoruba, Kongo, Akan, among others)—forged a shared identity rooted in Protestant Christianity, the English language, and U.S.-style republican institutions (Ciment, 2013; Nevin, 2011; Gershonim, 1985). Another instructive case is the Mormon migration to Utah, where religious refugees fleeing persecution established cooperative settlements and communal infrastructure that laid the foundations of statehood.[38]

Notwithstanding the particularities of our setting, its implications may be useful to other contexts. The making and functioning of local communities in Mandatory Palestine influenced the broader process of nation-building, though in ways often overlooked by the economic literature. Crucially, this effect passed through the nurturing of leaders:

[38]Historians have noted both parallels and contrasts between Zionist and Mormon settlement strategies (see, e.g., Arrington (2005) on the role of religious identity and communal institutions in Mormon cooperative settlements; Katz and Lehr (1991) on divergences between Mormon and Jewish agricultural experiments in western Canada).

while "pro-active" communities may not have succeeded in exporting their institutional model to the nation as a whole, they did export their knowledge and practices through the emergence of societal leadership.

References

Abramitzky, Ran, "The limits of equality: Insights from the Israeli kibbutz," *The Quarterly Journal of Economics*, 2008, *123* (3), 1111–1159.

_ , *The mystery of the Kibbutz: egalitarian principles in a capitalist world*, Vol. 73, Princeton University Press, 2018.

_ **and Victor Lavy**, "How responsive is investment in schooling to changes in redistributive policies and in returns?," *Econometrica*, 2014, *82* (4), 1241–1272.

Abu-Sitta, Salman H, *The Palestinian Nakba, 1948: The register of depopulated localities in Palestine*, Palestinian Return Centre, 2000.

Alan, Sule, Ceren Baysan, Mert Gumren, and Elif Kubilay, "Building social cohesion in ethnically mixed schools: An intervention on perspective taking," *The Quarterly Journal of Economics*, 2021, *136* (4), 2147–2194.

Alesina, Alberto and Eliana La Ferrara, "Ethnic diversity and economic performance," *Journal of Economic Literature*, 2005, *43* (3), 762–800.

Alff, Kristen, "Changing Capitalist Structures and Settler-Colonial Land Purchases in Northern Palestine, 1897–1922," *International Journal of Middle East Studies*, 2023, *55* (4), 675–692.

Allport, Gordon Willard, *The nature of prejudice*, Cambridge University Press, 1954.

Arrington, Leonard J, *Great basin kingdom: An economic history of the Latter-day Saints, 1830-1900*, Harvard University Press, 2005.

Assouad, Lydia, "Charismatic leaders and nation building," Technical Report 2020.

Azoulay, Ariella, "Civil alliances–Palestine, 1947–1948," *Settler Colonial Studies*, 2014, *4* (4), 413–433.

Bachi, Roberto, *The population of Israel*, Vol. 11, CICRED, 1974.

Bai, Ying, Ruixue Jia, and Jiaojiao Yang, "Web of Power: How Elite Networks Shaped War and Politics in China," *The Quarterly Journal of Economics*, 2022.

Bandiera, Oriana, Myra Mohnen, Imran Rasul, and Martina Viarengo, "Nation-building through compulsory schooling during the age of mass migration," *The Economic Journal*, 2019, *129* (617), 62–109.

Bazzi, Samuel, Arya Gaduh, Alexander D Rothenberg, and Maisy Wong, "Unity in diversity? How intergroup contact can foster nation building," *American Economic Review*, 2019, *109* (11), 3978–4025.

_ , **Martin Fiszbein, and Mesay Gebresilasse**, "Frontier culture: The roots and persistence of "rugged individualism" in the United States," *Econometrica*, 2020, *88* (6), 2329–2368.

Ben-Bassat, Yuval, "The Challenges Facing the First Aliyah Sephardic Ottoman Colonists," *Journal of Israeli History*, 2016, *35* (1), 3–15.

Billings, Stephen B, Eric Chyn, and Kareem Haggag, "The long-run effects of school racial diversity on political identity," *American Economic Review: Insights*, 2021, *3* (3), 267–84.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel, "Quasi-experimental shift-share research designs," *The Review of Economic Studies*, 2022, *89* (1), 181–213.

Brown, Jacob R, Ryan D Enos, James Feigenbaum, and Soumyajit Mazumder, "Childhood cross-ethnic exposure predicts political behavior seven decades later: Evidence from linked administrative data," *Science Advances*, 2021, *7* (24), eabe8432.

Buggle, Johannes, Thierry Mayer, Seyhun Orcan Sakalli, and Mathias Thoenig, "The Refugee's Dilemma: Evidence from Jewish Migration out of Nazi Germany," *The Quarterly Journal of Economics*, 2023, *138* (2), 1273–1345.

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

Carlitz, Ruth, Ameet Morjaria, Joris Mueller, and Philip Osafo-Kwaako, "State Building in a Diverse Society," *Review of Economic Studies*, 2024.

Chriqui, Avihay and Inbal Yahav, "HeBERT & HebEMO: a Hebrew BERT Model and a Tool for Polarity Analysis and Emotion Recognition," *INFORMS Journal on Data Science*, 2022.

Ciment, James, *Another America: The story of Liberia and the former slaves who ruled it*, Macmillan, 2013.

Clingingsmith, David, Asim Ijaz Khwaja, and Michael Kremer, "Estimating the impact of the Hajj: religion and tolerance in Islam's global gathering," *The Quarterly Journal of Economics*, 2009, *124* (3), 1133–1170.

Diao, Liang, "Collective defense by common property: The rise and fall of the Kibbutz." PhD dissertation 2021.

Diman, Amiad Haran, "My Neighbor, My Friend? The Logic of Intercommunal Cooperation During Armed Conflict," Technical Report 2024.

Dippel, Christian, "Forced coexistence and economic development: evidence from Native American Reservations," *Econometrica*, 2014, *82* (6), 2131–2165.

_ **and Stephan Heblich**, "Leadership in Social Movements: Evidence from the"Forty-Eighters" in the Civil War," *American Economic Review*, 2021, *111* (2), 472–505.

Duncan, Brian and Stephen J Trejo, "Intermarriage and the intergenerational transmission of ethnic identity and human capital for Mexican Americans," *Journal of Labor Economics*, 2011, *29* (2), 195–227.

Eldin, Munir Fakher, "Confronting a colonial rule of property," *The Arab Studies Journal*, 2019, *27* (1), 12–33.

Engel, David, "Zionism and the Negation of the Diaspora," in Hasia R Diner, ed., *The Oxford Handbook of the Jewish Diaspora*, Oxford University Press, 2021, pp. 151–164.

Fouka, Vasiliki, Soumyajit Mazumder, and Marco Tabellini, "From immigrants to Americans: Race and assimilation during the Great Migration," *The Review of Economic Studies*, 2022, *89* (2), 811–842.

Gelber, Yoav, *Palestine 1948: War, escape and the emergence of the Palestinian refugee problem*, Liverpool University Press, 2006.

Gershonim, Yekutiel, *Black Colonialism, the Americo-Liberian Scramble for the Hinterland*, Boulder: Westview Press, 1985.

Giuliano, Paola and Marco Tabellini, "The seeds of ideology: Historical immigration and political preferences in the United States," Technical Report, National Bureau of Economic Research 2020.

Goldberg, H.E., "Historical and Cultural Dimensions of Ethnic Phenomena in Israel," in Alex Weingrod, ed., *Studies in Israeli Ethnicity: After the Ingathering*, New York: Taylor & Francis, 1985, chapter 9, pp. 179–200.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift, "Bartik instruments: What, when, why, and how," *American Economic Review*, 2020, *110* (8), 2586–2624.

Gorny, Yosef, "The "Melting Pot" in Zionist Thought," *Israel Studies*, 2001, *6* (3), 54–70.

Greif, Avner, "Contract enforceability and economic institutions in early trade: The Maghribi traders' coalition," *American Economic Review*, 1993, *83* (3), 525–548.

Grootendorst, Maarten, "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," *arXiv preprint arXiv:2203.05794*, 2022.

Habyarimana, James, Macartan Humphreys, Daniel N. Posner, and Jeremy M. Weinstein, "Why does ethnic diversity undermine public goods provision?," *American Political Science Review*, 2007, *101* (4), 709–725.

Hacohen, Dvora, *Immigrants in Turmoil: Mass Immigration to Israel and its Repercussions in the 1950s and After*, Syracuse University Press, 2003.

Hasson, Shlomo and N Gosenfield, "Israeli frontier settlements: a cross-temporal analysis," *Geoforum*, 1980, *11* (4), 315–334.

Hengl, Tomislav, "Soil bulk density (fine earth) 10 x kg / m-cubic at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution," 2018.

_ **and Ichsani Wheeler**, "Soil organic carbon content in x 5 g / kg at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution," 2018.

Hijmans, Robert J, Susan E Cameron, Juan L Parra, Peter G Jones, and Andy Jarvis, "Very high resolution interpolated climate surfaces for global land areas," *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 2005, *25* (15), 1965–1978.

Hogan, Matthew, "The 1948 massacre at Deir Yassin revisited," *The Historian*, 2000, *63* (2), 309–334.

Horowitz, Dan and Moshe Lissak, *Trouble in utopia: The overburdened polity of Israel*, Suny Press, 2012.

_ **,** _ **, and Charles Hoffman**, *Origins of the Israeli polity: Palestine under the mandate*, University of Chicago Press, 1978.

Kafkafi, Eyal, "Segregation or integration of the Israeli Arabs: two concepts in Mapai," *International Journal of Middle East Studies*, 1998, *30* (3), 347–367.

Katz, Yossi and John C Lehr, "Jewish and Mormon agricultural settlement in Western Canada: a comparative analysis," *Canadian Geographer/Le Géographe Canadien*, 1991, *35* (2), 128–142.

Khalidi, Rashid, *Palestinian identity: The construction of modern national consciousness*, Columbia University Press, 1997.

Khalidi, Walid, "Plan Dalet: Master plan for the conquest of Palestine," *Journal of Palestine Studies*, 1988, *18* (1), 4–19.

_ , *All That Remains: The Palestinian Villages Occupied and Depopulated by Israel in 1948*, Institute for Palestine Studies, 2008.

Kimmerling, Baruch, "The Formation Process of Palestinian Collective Identities: The Ottoman and Colonial Periods," in Baruch Kimmerling, ed., *Clash of identities: Explorations in Israeli and Palestinian societies*, Columbia University Press, 2008, chapter 3, pp. 58–84.

Kranz, Dani, "Changing definitions of Germanness across three generations of Yekkes in Palestine/Israel," *German Studies Review*, 2016, *39* (1), 99–120.

Lazear, Edward P, "Culture and language," *Journal of Political Economy*, 1999, *107* (S6), S95–S126.

Masalha, Nur, "Expulsion of the Palestinians," *Washington, DC: Institute for Palestine Studies*, 1992.

Michalopoulos, Stelios and Elias Papaioannou, "The long-run effects of the scramble for Africa," *American Economic Review*, 2016, *106* (7), 1802–48.

Miguel, Edward and Mary Kay Gugerty, "Ethnic diversity, social sanctions, and public goods in Kenya," *Journal of Public Economics*, 2005, *89* (11-12), 2325–2368.

Morris, Benny, *The birth of the Palestinian refugee problem revisited*, Cambridge University Press, 2004.

Muchlinski, David, *Swords and Plowshares Jewish non-Weberian Governance in British Palestine*, Phd dissertation, Arizona State University, 2013.

Müller, Ulrich K and Mark W Watson, "Spatial correlation robust inference," *Econometrica*, 2022, *90* (6), 2901–2935.

Near, Henri, *The Kibbutz Movement: A History. Origins and Growth, 1909–1939*, Oxford University Press, 1992.

Near, Henry, *Where community happens: the kibbutz and the philosophy of communalism*, Peter Lang International Academic Publishers, 2011.

Nevin, Timothy D., "The Uncontrollable Force: A Brief History of the Liberian Frontier Force, 1908–1944," *The International Journal of African Historical Studies*, 2011, *44* (2), 275–297.

Panza, Laura and Eik Swee, "Fanning the flames: Rainfall shocks, inter-ethnic income inequality, and conflict intensification in Mandate Palestine," *Journal of Economic Behavior & Organization*, 2023, *206*, 71–94.

Pappe, Ilan, *La pulizia etnica della Palestina*, Fazi Editore, 2015.

Penslar, Derek Jonathan, "Technical expertise and the construction of the rural Yishuv, 1882–1948," *Jewish History*, 2000, *14* (2), 201–224.

Pettigrew, Thomas F and Linda R Tropp, "A meta-analytic test of intergroup contact theory," *Journal of Personality and Social Psychology*, 2006, *90* (5), 751.

Rabin, Albert Í, *Growing up in the kibbutz*, Springer, 2013.

Rayman, Paula M, *Kibbutz Community and Nation Building*, Vol. 562, Princeton University Press, 1981.

Reichman, Shalom and Shlomo Hasson, "A cross-cultural diffusion of colonization: From Posen to Palestine," *Annals of the Association of American Geographers*, 1984, *74* (1), 57–70.

Rohner, Dominic and Ekaterina Zhuravskaya, "The Economics of Nation-Building: Methodological Tool Kit and Policy Lessons," *Annual Review of Economics*, 2024, *17*.

Rozin, Orit, "State and Society Building in Early Israel," in Reuven Y Hazan, Alan Dowty, Menachem Hofnung, and Gidon Rahat, eds., *The Oxford Handbook of Israeli Politics and Society*, 2018.

Sabbagh-Khoury, Areej, *Colonizing Palestine: The Zionist left and the making of the Palestinian Nakba*, Stanford University Press, 2023.

Sanderson, Eleanor and Frank Windmeijer, "A weak instrument F-test in linear IV models with multiple endogenous variables," *Journal of econometrics*, 2016, *190* (2), 212–221.

Sela-Sheffy, Rakefet, "Integration through Distinction: German-Jewish Immigrants, the Legal Profession and Patterns of Bourgeois Culture in British-ruled Jewish Palestine 1," *Journal of Historical Sociology*, 2006, *19* (1), 34–59.

Sellers, Jefferey M, Anders Lidström, and Yooil Bae, *Multilevel democracy: How local institutions and civil society shape the modern state*, Cambridge University Press, 2020.

Shafir, Gershon, *Land, labor and the origins of the Israeli-Palestinian conflict, 1882-1914*, Univ of California Press, 1996.

Shalumov, Vitaly and Harel Haskey, "HeRo: RoBERTa and Longformer Hebrew Language Models," *arXiv:2304.11077*, 2023.

Smooha, Sammy, "The Jewish Ethnic Divide and Ethnic Politics in Israel," in Reuven Y Hazan, Alan Dowty, Menachem Hofnung, and Gidon Rahat, eds., *The Oxford Handbook of Israeli Politics and Society*, 2018.

Spiro, Melford E, *Kibbutz; venture in Utopia.*, Harvard University Press, 1956.

Stein, Kenneth W, "The Jewish National Fund: land purchase methods and priorities, 1924–1939," *Middle Eastern Studies*, 1984, *20* (2), 190–205.

Sternhell, Zeev, *The founding myths of Israel: Nationalism, socialism, and the making of the Jewish state*, Princeton University Press, 2009.

Tidhar, David, *Encyclopedia of the Founders and Builders of Israel*, Tel Aviv, 1947.

Trezib, Joachim and Ines Sonder, "The Rassco and the Settlement of the Fifth Aliyah: Pre-State and Early State Middle Class Settlement and its Relevance for Public Housing in Eretz-Israel," *Israel Studies*, 2019, *24* (1), 1–23.

Weingrod, Alex, *Studies in Israeli Ethnicity: After the Ingathering*, Taylor & Francis, 1985.

Weintraub, Dov, M Lissak, and Y Azmon, *Moshava, Kibbutz, Moshav: Jewish Settlement and Colonization in Palestine*, Cornell University Press, 1969.

Ya'ar, Ephraim, "Continuity and change in Israeli society: The test of the melting pot," *Israel Studies*, 2005, *10* (2), 91–128.

Yitzhaki, Shlomo and Edna Schechtman, "The "melting pot": A success story?," *The Journal of Economic Inequality*, 2009, *7* (2), 137–151.

Yona, Rona, "A kibbutz in the diaspora: The pioneer movement in Poland and the Klosova kibbutz," *Journal of Israeli History*, 2012, *31* (1), 9–43.

ONLINE APPENDIX—not for publication

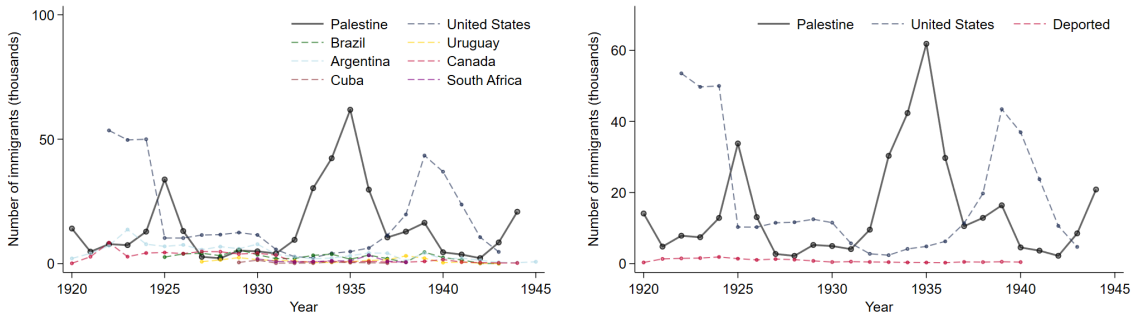
A Data appendix

This section provides complements to the data description of Section 2 with: a detailed presentation of our data sources (except our data sources for leaders, rather described in Appendix C); and additional evidence about our "frontier expansion" algorithm.

A.1 Data sources

This project relies on numerous data sources that are briefly described in Section 2 and more thoroughly presented below.

Figure A1. Jewish migration across destinations (1920–1945).



(a) All destinations (b) United States

Notes: Panel (a) displays migration (in thousand migrants) to Palestine (black, plain line), the United States (blue, dashed), Brazil (green, dashed), Uruguay (yellow, dashed), Argentina (light blue, dashed), Canada (red, dashed), Cuba (maroon, dashed), and South Africa (purple, dashed). Panel (b) shows the (limited) number of Jews deported from the United States or debarred entry (red, dashed).

Migration Our migration statistics come from The American Jewish Year Book compiled by The Bureau of Jewish Social Research.[39] These data sources also contain statistics about: migration to other destinations (more specifically, the United States); deportations from the United States; emigration of Jews from Mandatory Palestine; and immigration of non-Jews into Mandatory Palestine.

In this Appendix, we leverage these data to show the importance of push factors from Europe and the substitution across the two preferred migration destinations, i.e., the United States and Mandatory Palestine. In Figure A1, we report Jewish migration across

[39]The aggregate statistics over the period were compiled by The Bureau of Jewish Social Research and published in The American Jewish Year Book from 1920–1921 to 1947–1948. These publications were mostly produced by Harry Sebee Linfield who reported on the Jewish community in the United States, but also compiled statistics from various sources including the timing and characteristics of settlers in Palestine. The micro-data is available within the Card catalog of the Statistics Bureau of the Jewish Agency's Immigration Department, 1919–1948.

various destinations between 1920–1945. Panel (a) displays migration (in thousand migrants) to Palestine (black), the United States (blue), Brazil (green), Uruguay (yellow), Argentina (light blue), Canada (red), Cuba (maroon), and South Africa (purple). One can see that migrants essentially target Mandatory Palestine and the United States, except during the end of the 1920s where: migration flows to the United States dry out (following the Immigration Act of 1924); and migration to Argentina, Brazil or Canada is then non-negligible. This figure also illustrates the importance of push factors: the aftermath of pogroms (e.g., during the Russian Civil War) in the early 1920s, the arrival of the Nazi party to power in Germany, or the start of the Second World War all produce large migration flows. Panel (b) focuses on the two main destinations and shows how they appear to be negatively correlated. Indeed, while demand for emigration (push factors) generates positive correlation in migrant flows across destinations, pull factors at those destinations do generate some substitution: for example, the Immigration Act of 1924 in the United States generated a rise in migration to Palestine; and the White Paper of 1939 limiting migration to Palestine redirected migrant flows to the United States.[40]

Figure A2. Emigration and (non-)Jewish migration to Mandatory Palestine.

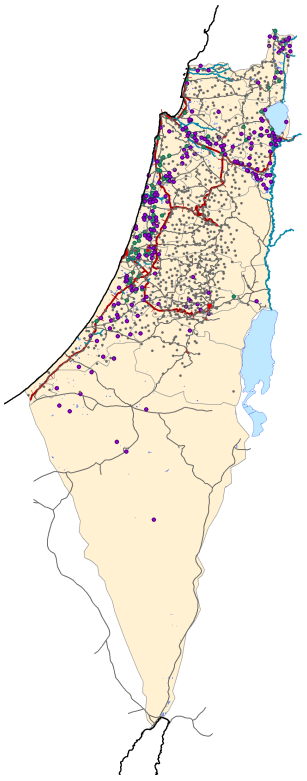
Notes: This Figure displays two ratios characterizing migration in Mandatory Palestine: (i) the ratio of the number of Jewish immigrants over the number of Jewish immigrants and the number of Jewish emigrants (a ratio capturing the relative importance of inflows versus outflows, in green); and (ii) the ratio of Jewish immigrants over the total number of immigrants (in red).

Finally, we shed some light on two dimensions that we do not discuss explicitly in Section 1.2: emigration from Mandatory Palestine and (non-)Jewish migration to Mandatory Palestine. Figure A2 shows that (non-)Jewish migration to Mandatory Palestine is limited—or reciprocally, almost all migrants over the period 1920–1945 are Jewish (red line) except during times of limited migration (1927–1932). Figure A2 also shows an immigration ratio for Jews, i.e., the ratio of the number of Jewish immigrants over the

[40]Note that this substitution does not usually arise from ex-post deportations, but from an ex-ante deterrence effect. We illustrate the limited number of Jews refused entry to the United States in panel (b) of Figure A1.

number of Jewish immigrants and the number of Jewish emigrants. In short, this ratio would be 0.50 if outflows were equal to inflows: the ratio is closer to 0.80-0.90 during times of large immigrant flows—in fact, the emigration rate is quite stable such that fluctuations only arise from the volatility in the immigration rate.

Figure A3. Geography of Palestine under British mandate.

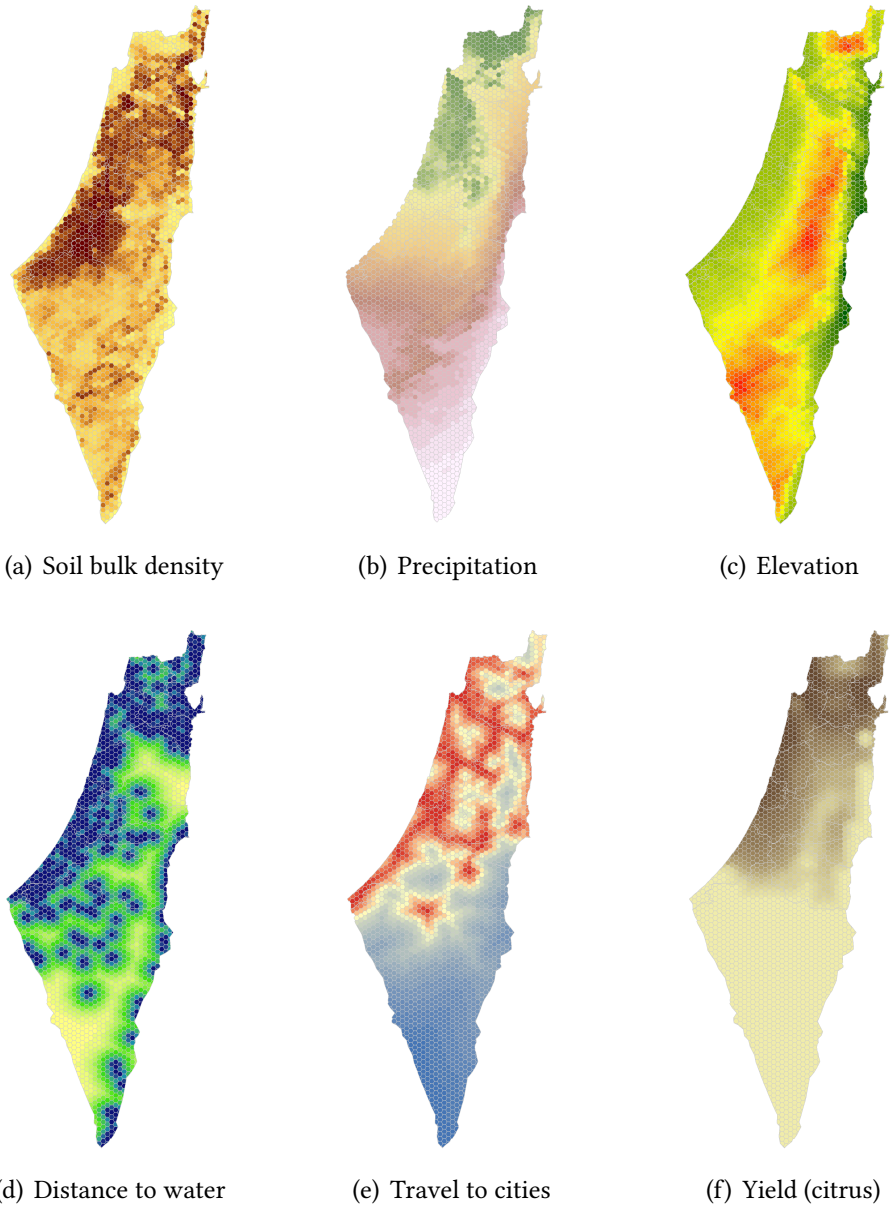


Notes: This Figure displays the geography of Palestine under British mandate. The dots indicate Arab settlements (gray) and Jewish settlements (green: agricultural; purple: mainly owned by the Histadrut—the General Organization of Workers in Israel; red: cities). Roads, railways and waterways as of 1920 are indicated with gray, red and blue lines respectively.

The geography of Mandatory Palestine To discipline our settlement creation algorithm, we divide the territory covered by Palestine under the British mandate in 2,698 hexagons of about 2-kilometer radius. The purpose is to nest all geographic variables at a very disaggregated level, and to compensate for the lack of relevant administrative units covering the territory beyond historical subdistricts. The implied spatial variation is shown in Figures 2 and 4 of the paper. While most of our empirical specifications are at the level of a settlement, we do hinge on the hexagon variation in the prediction procedure, e.g., as reported in Table 1 and Appendix Tables A3 and A2.

We extract high-quality data on topography (30m precision), soil quality (100m precision) and temperature/precipitation (1km precision) from Google Earth Engine. More specifically, we construct the maximum, minimum and average elevation within each

Figure A4. The geography of Mandatory Palestine.



(a) Soil bulk density (b) Precipitation (c) Elevation

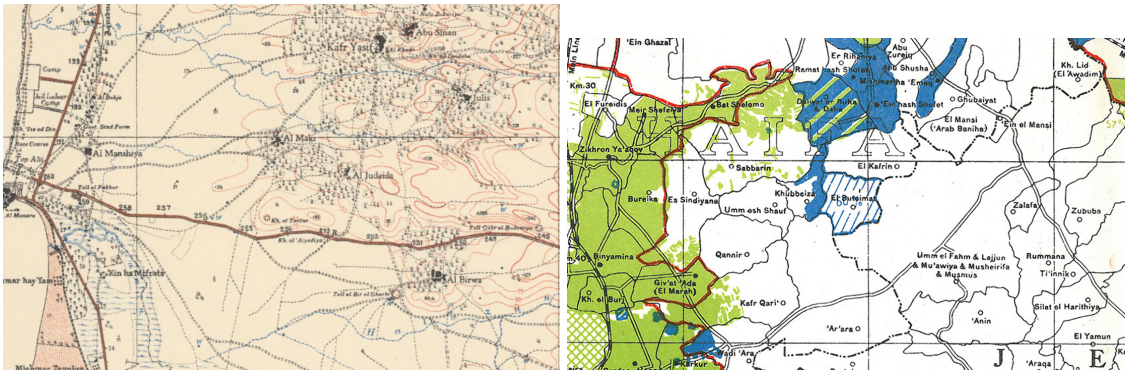
(d) Distance to water (e) Travel to cities (f) Yield (citrus)

Notes: The main geographic unit is a hexagon with a radius of 2 kilometers. The color code is as follows: yellow to dark brown from low to high bulk density in panel (a); white to red then yellow and green from low to higher precipitation in panel (b); green to red from low to high elevation in panel (c); blue to green and then yellow from higher to lower distance to fresh water in panel (d); blue to dark red from high to low travel time to nearest large city in panel (e); yellow to brown from low to high yield for citrus (intermediate rain scenario, source: GAEZ/FAO) in panel (f).

hexagon; the average slope; the soil bulk density at 0 cm depth as reconstructed from recent satellite imagery (Hengl, 2018); the organic carbon content, also inferred from recent satellite imagery (Hengl and Wheeler, 2018); and the average annual precipitation and temperature (Hijmans et al., 2005). We construct the distance to the nearest shore, river, canal, stream, wadi, and the length and area of water bodies. Finally, we

gather data on the potential yield per hectare for various crops (citrus in particular) as extracted from the Global Agro-Ecological Zones Agricultural Suitability and Potential Yields (GAEZ). We represent a subsample of these characteristics in Figure A4—nested at the level of hexagons. These characteristics are used in our "frontier expansion" algorithm presented in Section 3.1 and as separate controls in our main empirical strategy (Section 3.2).

Figure A5. Transportation and land ownership in Mandatory Palestine.



(a) Survey of Palestine (b) Index to Villages and Settlements

Sources: Survey of Palestine, 1935–1938, east of Acre, in panel (a); Index to Villages and Settlements, 1944, Haifa sub-district.

We retrieve the transportation infrastructure from early maps provided as part of the Survey of Palestine (1917–1918), which we calibrate against the later, better waves (Survey of Palestine, 1935–1938, see an example in Figure A5, panel a). We recover main roads, secondary roads, railways, navigable rivers, and canals. We then collapse the data at the hexagon level by computing: the distance to and length of roads and railways as of 1920; the distance and travel time to the nearest existing city and the distance and travel time to Haifa, Jerusalem and Tel-Aviv/Jaffa; and the distance/travel time to existing Arab settlements. As for the previous geographic controls, these transportation/market access controls are both used in our "frontier expansion" algorithm presented in Section 3.1 and as separate controls in our main empirical strategy (Section 3.2).

We further extract land shares in 1881 across 10 categories (built-up, woodland, gardens, wetlands, orchards, dunes, open forests, scrub, vineyards, alluvial sands) as extracted from processed maps provided by the Palestine Exploration Fund, which we use in a robustness check.

We digitize the maps available within the Index to Villages and Settlements (1944, see Figure A5, panel b), reporting the local land claimed by Jewish companies and private owners (full, concession, shares in undivided land) and by the Jewish National Fund (full, concession, shares in undivided land); and later land regulations imposed by the British

mandate ("zone A", "zone B", "free zone"). We exploit some of these land characteristics in our "frontier expansion" algorithm (Section 3.1).

Pacts, destruction of Arab and Jewish localities, and civilian casualties We rely on the efforts of historians in order to construct proxies for inter-ethnic relationships prior to and during the conflict provoked by the Partition Plan for Palestine. We first use data collected in the Haganah archives by Azoulay (2014) on 100 pacts between Arab and Jewish communities and meetings "during which the participants raised demands, sought compromises, set rules, formulated agreements, made promises, sought forgiveness, and made efforts to compensate and reconcile". This data, probably selective in that other such meetings could have taken place without being recorded, is also used in recent research by Haran Diman (2024). Second, we reconstruct the 1947–1949 civil war at the village level from various data sources. We mostly draw on the Atlas of Palestine which records hundreds of attacks, civilian casualties and depopulated Arab villages during 1947–1948 (based on various sources). We also capture whether the local militias defended the village, drawing on Abu-Sitta (2000) and Khalidi (2008), which we use as a control variable in all our specifications looking at civilian casualties.

A.2 Meeting minutes

Data sources Our primary data sources are the Yad Tabenkin archives (covering the United Kibbutz Movement, HaKibbutz HaMeuhad and Chever HaKvutzot). These archives gather meeting notes from the following meeting locations (with the dates in parentheses): Tel Aviv (1930, 1931, 1937, 1940), Givat Hashlosha (1932, 1943), Hadera (1930, 1933, 1936), Degania A (1933, 1935, 1945), Degania B (1934), Dalhamia/Gesher (1935), Ganigar (1935, 1942), Ein Harod (1936, 1946), Geva (1937), Beit Hashita (1937), Mazkirut HaPeila (1937), Gedera (1938), Na'an (1938), Givat Brenner (1939), Ashdot Yaakov (1940), Revivim (1939, 1940, 1943), Kfar Saba (1940), Hulda (1940, 1941), Ein Gev (1943), Ein Hayam (1943, 1945), Ramat David (1943), Beit Oren (1944), Alonim (1944, 1946), Shafayim (1945, 1946), Haifa (1947). Within these minutes, the settlements/cities that are most frequently mentioned are: Afikim, Alonim, Bat Yam, Beit Alpha, Beit Hanan, Beit Hashita, Beit Oren, Caesarea, Dalhamia, Degania A, Degania B, Eilat, Ein Ganim, Ein Gev, Ein Harod, Ein Hayam, Ganigar, Gat Rimon, Gesher, Geva, Gevat, Givat Brenner, Givat Haim, Givat Hashlosha, Gordonia, Hadera, Hafetz Haim, Haifa, HaShachar, Herzliya, Hulda, Kfar Giladi, Kfar HaSharon, Kfar Mefia, Kfar Saba, Kfar Yizrael, Kinneret, Khirbet Zababdeh, Nahalat, Ness Ziona, Mazkirut HaPeila, Menahemia, Merhavia, Mishmar HaYam, Na'an, Petah Tikva, Ra'anana, Rama HaKovesh, Ramat David, Ramat Rachel, Rehovot, Revivim, Rishon LeZion, Shfayim, Tel Aviv, Tel Yosef, Yehuda, Yagur, Yigur. In effect, the textual

analysis—which we describe below—identifies about 120 settlements of interest. We also gather documents from the Moshav Movement Records stored at The National Library of Israel or Israel State Archives.

Data construction These minutes are unstructured in many aspects: (i) they report oral discussions with different participants from different background, (ii) the subject can be specific (about one settlement) or general (about organization or geopolitical issues), and (iii) the minutes use abbreviations and unclear references to previous discussions. In addition, the transcription from Optical Character Recognition might introduce noise. We process these minutes as follows. In a first step, we digitize them using a mixture of Optical Character Recognition and manual transcription (e.g., for handwritten notes and poorly-scanned documents). In a second step, we use two parallel procedures to correct the text and translate it to English (DeepL Translate; and a large language model, GPT-4o). In a third step, we produce a topic analysis in a large language model (GPT-4o), which we probe by using a transformer-based topic modeling based on BERT and applied to the translated minutes (BERTopic, see Grootendorst, 2022) and a similar topic analysis based on BERT models developed and adapted to Hebrew by Chriqui and Yahav (2022) and Shalumov and Haskey (2023). We report below our full API-based prompt:

```
file_names = os.listdir(path)
papers = []
for file_name in file_names:
  if file_name.endswith('.txt'):
    with open(path + '/' + file_name, "r", encoding="utf-8") as file:
        fulltext = file.read()
        fulltext = fulltext.split("\n")
        print("2.a topics:",file_name,len(fulltext))
        N = paragraphs
        n = math.floor(len(fulltext)/N) + 1
        translated_text = []
        for x in range(n):
            text = fulltext[x*N:(x+1)*N-1]
            text = '\n'.join(text)
            prompt = f'''Below are the minutes of meetings organized in communities of Mandatory Palestine between 1925 and 1947. Please identify the communities, kibbutzim or locations mentioned in these minutes. For each of the communities that you have identified, provide: (i) the topics that are discussed, (ii) whether they are of major or minor importance from 100 (very important) to 0 (not important), (iii) the tone in which they are mentioned, and (iv) the main quotes justifying why this topic is included. Only provide an entry if there is convincing material to support your choice.

            Note that, when a location is not specified, the default location should be {file_name[:-4]} if the latter is a location or the initial location mentioned in the minutes.

            The topics should be selected among the following categories: land, housing, water, agriculture, infrastructure, community cohesion, finance, economy, security, absorption of immigration, language, recruitment, army, secretariat, ideology, youth, education.

            The quotes should have an explicit reference to the chosen topic.

            The tone should be selected among the following categories: urgent, worried, critical, neutral, optimistic, constructive.

            The output should be a JSON list with the following format
            [
                {{"community": "<community1>", "topic": "<topic1>", "importance": "<importance>", "tone": "<tone>, "quote": "<quote1>"}},
```

```
                {{"community": "<community1>", "topic": "<topic2>", "importance": "<importance>", "tone": "<tone>", "quote": "<quote2>"}},
                {{"community": "<community1>", "topic": "<topic3>", "importance": "<importance>", "tone": "<tone>", "quote": "<quote3>"}},
                {{"community": "<community1>", "topic": "<topic4>", "importance": "<importance>", "tone": "<tone>", "quote": "<quote4>"}},
                ...
            ]

            Please only report the JSON list. You can find the minutes below.


            '''
            completion = client.chat.completions.create(
                model = "gpt-4o",
                temperature = 0,
                messages = [
                    {"role": "system", "content": "You are a helpful historian specialized in Mandatory Palestine"},
                    {"role": "user", "content": prompt + text}
                    ]
                )
            topics = completion.choices[0].message.content
            topics = topics.replace("```", "")
            topics = topics.replace("json", "")
            if x == 0:
                topics_name = file_name[:-4] + '.json'
            else:
                topics_name = file_name[:-4]+'_p' + str(x) + '.json'
            with open(path_save + '/' + topics_name, "w", encoding="utf-8") as topics_file:
                topics_file.write(topics)
                topics_file.close()
```

The output from the previous procedure is much richer and detailed than our unsupervised topic modeling approaches, such that we only use the latter in unreported checks (a previous version of the paper was based on these alternative approaches). In particular, it allows us to classify topics across numerous categories (land, housing, water, agriculture, infrastructure, community cohesion, finance, economy, security, absorption of immigration, language, recruitment, army, secretariat, ideology, youth, education) and to qualify the tone of discussions (urgent, worried, critical, neutral, optimistic, constructive). We provide below a few examples of the extracted discussions for illustration. In Alonim, discussants were worried about (the lack of) immigration: “Alonim has many possibilities. It is our duty to exploit them as much as we can. The lack of people is a major hindrance.” In Yigur, the discussions were about education and ideology: “We need to discuss this, how we educate our friends in the diaspora and what will be the way of the movement.” In Aiyanot (Ramat David), discussions were about the financial state of the kibbutz: “We entered into a heavy burden of debts, a severe financial burden, which I will discuss later.” Meeting notes were also discussing security concerns: “And recently we have suffered another loss, Tzilla Cohen was murdered on the roads of the land”, or “Three members from Hulda have served their full sentence. They were granted no clemency, no mitigation of sentence by the military commander, which is usually granted to every criminal, even to every Arab criminal.” In Kfar Hahoresh, there were cohesion issues: “We brought in a group of young members who just disembarked from the ship to a place isolated from the surrounding movement, under working condi-

tions and a wonderful climate, to a situation of splitting and dispersal into several groups, without the possibility of teamwork and internal life cohesion of the group." Integrating members was also a challenge in Gordonia: "Many members lack any knowledge or understanding of the group's aspirations, organization, and public relations."

Additional information Finally, we extract the following additional information from these minutes using similar approaches as the one described above: (i) references to Arab neighbors, which we discuss in Section 4.2 and Figure 8; (ii) the allocation of migrants across settlements, which we use in a robustness check in Appendix B.1 and Figure B2; and (iii) the names of meeting participants for each of the hundreds of meetings covered by our minutes. We do not exploit the latter in the paper, because the number of identified, frequent participants who become "leaders" is limited. More specifically, we identify: Shulamit Aloni (Tel Aviv), Israel Bar-Yehuda (Yagur), David Ben-Gurion, Gideon Ben-Israel (Haifa), Yosef Efrati (Be'er Tuvia), David Elazar (Ein Shemer), Eliezer Kaplan (Tel Aviv), Aryeh Leib Frumkin (Petah Tikva), Heshel Frumkin (Deganiya B, Tiberias), Yisrael Galili (Naan), Yehoshua Hankin (Gedera), Pinhas Lavon (Hulda), Peretz Naftali (Tel Aviv), Haim-Moshe Shapira (Jerusalem), Yitzhak Tabenkin (Kinneret, Merhavia, Ein Harod), as frequent contributors. Their role in the Yishuv is well known. The other leaders are thus not present or sufficiently active in those meetings for us to perform a more systematic analysis. Reciprocally, a number of very active participants ended up not assuming responsibilities at the national level: Aba Ben-Yaakov, Shmuel Bernstein (Nahalat Yitzhak, killed in 1939), Y. Bitman (Kfar HaHoresh), Chiyuta Bussel (Petach Tikva), Reuven Vinya Cohen (Ein Harod), Yedidyah Freund (Hulda), Israel Galer (Alonim), Aharon Friedman, Benny Marshak (whose contribution to the kibbutz movement is described in Yona, 2012), M. Mendel (Kfar HaHoresh), Shalom Zak (Givat Hashlosha).

A.3 Photographs

Data sources To measure community cohesion and inter-community relations, we rely on the Eric and Edith Matson Photograph Collection available online from the Library of Congress. The Library of Congress has curated these resources produced by the American Colony Photo Department in Jerusalem (and its successor, the Matson Photo Service), such that each glass plate negative comes with a title, an estimated date, a description of the subject, a collection, a medium ("1 negative : nitrate ; 4 x 5 in." or "1 transparency : glass ; 5 x 7 in."), and—important to our purpose—a location either from the title or inferred by the Library of Congress. We pre-process the data through textual analysis of the title and an image recognition algorithm to extract the objects depicted

Figure A6. Photographs from the Eric and Edith Matson Photograph Collection—two examples.

(a) A match factory (b) Demonstrations

Notes: Panel (a) displays a photograph depicting a child working in a match factory around 1930 in Nablus. The original title is "Arab factories & gen[eral] improvements in Nablus. Match factory. Labeling match boxes by machines". Panel (b) displays a photograph of Arab demonstrations around the Jaffa gate in 1933. The original title of the collection is "Arab demonstrations on Oct. 13 and 27, 1933. In Jerusalem and Jaffa. Demonstrators facing police baton charge".

by each picture. We also harmonize location names. Based on this pre-processing, we select about 3,400 pictures which can be assigned to more than 80 settlements in total and 45 of those within our sample of settlements. Figure A6 provides two examples, one taken from an Arab factory in Nablus and one covering Arab riots in 1933.

Data construction We provide a textual and image input to a large language model (GPT-4o) in order to capture the emotions conveyed by each photograph (anger, amusement, awe, contentment, disgust, excitement, fear, sadness). The prompt is: "What is the emotion conveyed by this {photograph}? Your answer will assign a probability to the following emotions: anger, awe, contentment, sadness, disgust, fear, excitement, amusement. Your answer will have the following format: [anger: 'value1', awe: 'value2', contentment: 'value3',...]. If this is helpful, the title of the image is: {title}" where {title} is the title of the picture in the collection and {photograph} is the relatively low-resolution image provided by the Library of Congress. Figure A7 shows images with the highest emotion score across the different categories. Negative emotions such as anger, disgust, fear or sadness are associated with riots, protests; positive emotions are associated with friendly pictures taken from farming schools or sporting events. In practice, we will typically group negative emotions together to generate a negativity (probabilistic) score between 0 and 1. In general, these negative emotions remain in the minority: most pictures display a high contentment (e.g., positive daily life interactions) or awe score (e.g., landscape)—see Figure A8.

The previous analysis however ignores whether the emotions come from those ex-

Figure A7. Photographs from the Eric and Edith Matson Photograph Collection—emotions.

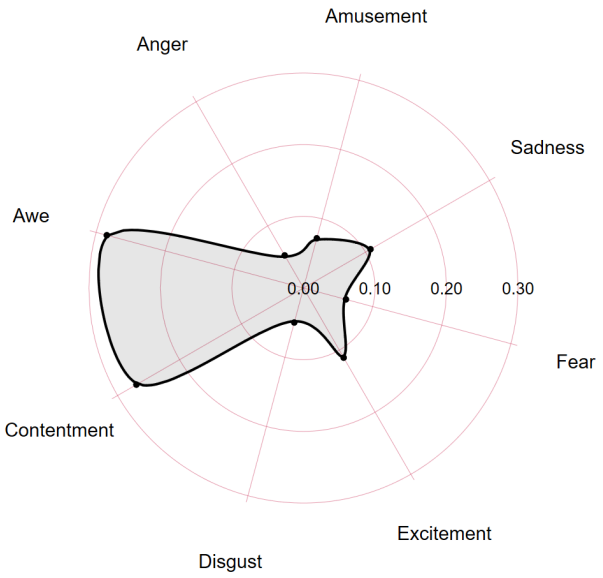
(a) Amusement (b) Anger (c) Awe (d) Contentment

(e) Disgust (f) Excitement (g) Fear (h) Sadness

Notes: These pictures display photographs associated with the highest emotion scores. For instance, the original title for panel (b) is: "Jewish anti Palestine White Paper demonstrations. [Women's demonstration on May 22, 1939]. Demonstration before the District Offices on the Jaffa Road, taken while one of the young women was emphatically denouncing the new polity". The original title for panel (d) is: "Nahalal. Girls' Agricultural Training School. Putting in tomato plants showing typical dress, head-scarf, middy, and bloomers". The original title for panel (e) is: "Palestine events. The 1929 riots, August 23 to 31. Jewish home plundered by Arab rioters in Hebron. Blood-stained floor littered with books".

pressed by the (human) subjects in the photographs. We thus focus on the nature of community relationships conveyed by the picture (hostile, amicable, neutral) with the following question: "This {scene} is from Mandatory Palestine between 1920 and 1948. Is this picture showing meaningful interactions between people? If so, your answer will assign a probability to the following adjectives characterizing these interactions: hostile, amicable, neutral. Your answer will have the following format: [hostile: 'value1', amicable: 'value2', neutral: 'value3']. If not, please answer: []. If this is helpful, the title of the image is: {title}". Figure A9 provides the highest amicable and hostile scores. As apparent, panel (a) was already covered in the previous picture; however, this more specific prompt allows us to limit our sample to 1,500 pictures with proper social interactions. Note that we found difficult to have a clearer view of whether community relations are within communities or between Arabs, Jews, and other groups (e.g., Druzes): the titles do not always provide this information, and the large language model does not provide such characterization with a sufficient accuracy (notwithstanding recent controversies leading to a recalibration of its answers to "ethnic questions").

Figure A8. Photographs from the Eric and Edith Matson Photograph Collection—the average emotion score.



Notes: This Figure displays the average emotion score (a probabilistic score) for 3,400 selected photographs which can be assigned to a settlement of our sample during our period of interest. The emotions are: anger, amusement, awe, contentment, disgust, excitement, fear, and sadness.

Figure A9. Photographs from the Eric and Edith Matson Photograph Collection—community relationships.

(a) Amicable (b) Hostile

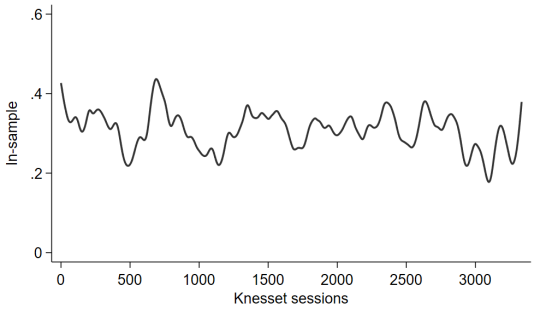
Notes: Panel (a) displays a photograph associated with the highest amicable score. The original title is: "The vintage season Zikh'ron Ya'aqov, July 24, 1939. Two girls pause to giggle & to sample the sweetness of the vine. Sampling grapes during a short pause". Panel (b) displays a photograph associated with a high hostility score. The original title of the collection is: "Result of terrorist acts & government measures. Russian police-woman searching Jewish female for arms at the Jaffa Gate".

A.4 Speeches

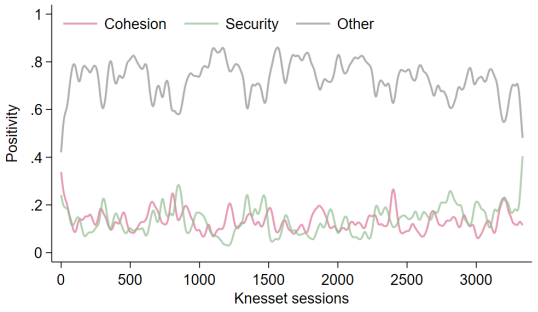
Data sources We compile approximately 3,340 minutes of Knesset sessions, covering the following legislative terms: 1st Knesset (1949–1951); 2nd (1951–1955); 3rd (1955–

1959); 4th (1959–1961); 5th (1961–1965); 6th (1965–1969); 7th (1969–1974); and 8th (1974–1977). Each document corresponds to a single session and includes records of all interventions, along with the names and official roles of the participants. Our textual analysis follows a methodology similar to that used for the previous meeting minutes, employing a large language model (GPT-4.1-mini) via an API prompt. A general prompt extracts, for each major intervention: (i) the speaker's name, (ii) the topic, (iii) a perceived importance score (ranging from 0 to 100), (iv) the tone, (v) a one-sentence summary, and (vi) a concise, descriptive title for the discussion (e.g., "Operation Susannah"). In addition, a targeted prompt identifies segments explicitly addressing Jewish-Arab relations.

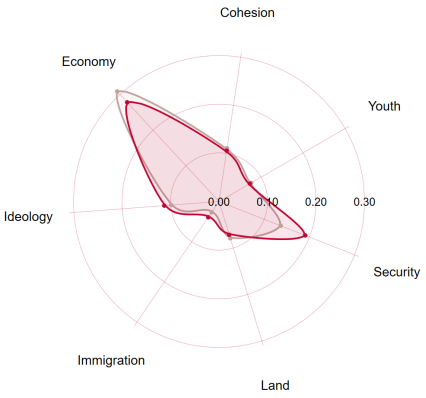
Figure A10. Speeches from the Knesset—a few descriptives.



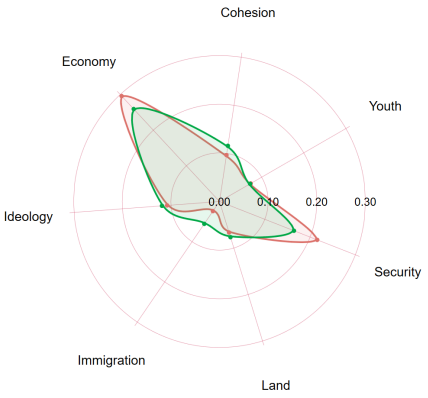
(a) Coverage



(b) Topics (trends)



(c) Topics (distribution)



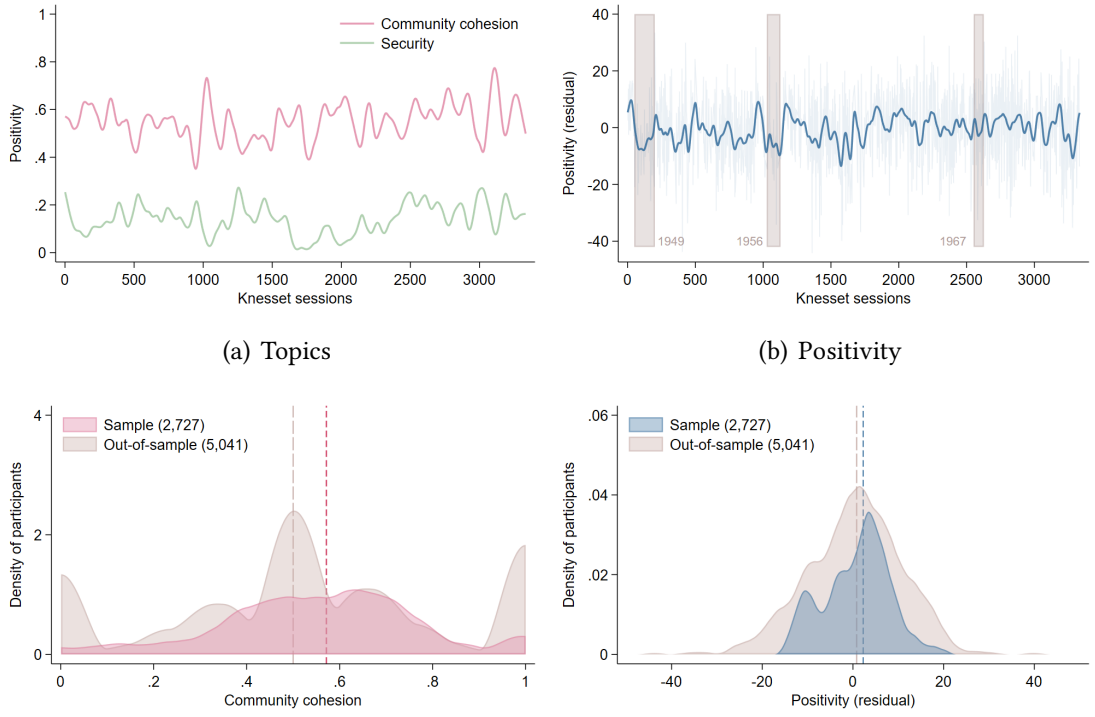
(d) Topics (diversity)

Notes: Panel (a) shows the incidence of interventions from leaders of our baseline sample across Knesset sessions between 1949–1977. Panel (b) shows the incidence of interventions about "community cohesion", "security", and "other". For visualization purposes, the series are smoothed using a Hodrick-Prescott filter. Panel (c) shows the distribution of interventions across selected topics for speakers within our sample (about 32% of all interventions, in red) and speakers outside of our sample. Panel (d) shows the distribution of interventions across selected topics for speakers within our sample: The red (green) curve shows the average share of entries for settlements with above-median (below-median) predicted measure of uniformity.

Descriptive statistics From the 3,340 session minutes, we identify 62,000 individual interventions, 25,420 distinct discussion subjects, and 20 consolidated thematic cate-

gories (e.g., immigration absorption, community cohesion, economy, infrastructure, security, public services, diplomacy).

Figure A11. Speeches from the Knesset about Arab-Jewish relationships—a few descriptives.



(a) Topics

(b) Positivity

(c) Topics (distr.)

(d) Positivity (distr.)

Notes: Panel (a) shows the incidence of interventions about Arab-Jewish relationships across Knesset sessions between 1949–1977, which respectively pertain to "community cohesion" and "security". For visualization purposes, we highlight major periods of conflicts, and the series are smoothed using a Hodrick-Prescott filter. Panel (b) shows a residualized measure of positivity, in which the measure is cleaned from topic indicators. Panel (c) shows the distribution of interventions along the probability to discuss community cohesion for speakers within our sample (2,727 interventions, about 35% of all interventions) and speakers outside of our sample (5,041 interventions). Panel (d) shows the distribution of interventions along their residualized positivity for speakers within our sample and speakers outside of our sample. In panels (c) and (d), the lines represent averages for the two groups of speakers.

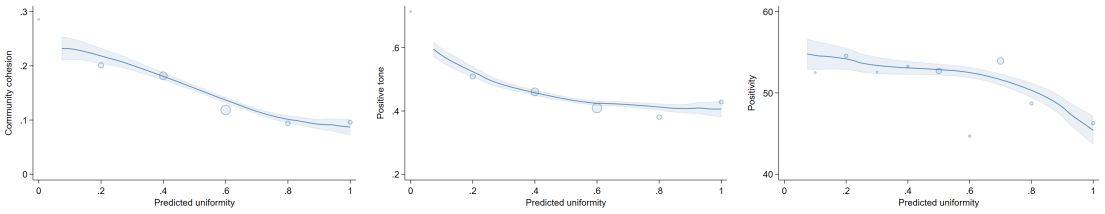
Figure A10 (a) shows the prevalence of leaders from our baseline sample across Knesset sessions—about 0.32 on average. Despite comprising only 120 individuals, these leaders appear frequently, possibly because they often steer political discourse within their respective parties. Figure A10 (b) indicates that although community cohesion and security are not among the most dominant topics, their discussion remains steady over time. In panel (c), we observe that in-sample leaders (in red, representing 32% of all interventions) are slightly more inclined to address immigration and security issues, and slightly less inclined to focus on economic matters compared to other speakers.[41] Fi-

[41]We categorize themes as follows: cohesion (community cohesion); absorption (immigration); ideology (including governmental issues); economy (including financial and infrastructure matters); land (housing, agriculture, water); security (including military affairs); and youth (education and language).

nally, Figure A10 (d) presents the distribution of topics among leaders based on whether they originate from settlements with above- or below-median predicted levels of uniformity: leaders from diverse settlements are more likely to discuss themes related to the absorption of immigration or community cohesion.

Figure A11 reports on mentions of Jewish-Arab relations. Panel (a) details the primary framing of these discussions (community cohesion against security matters), while panel (b) illustrates the tonal fluctuations over time. Of the 7,768 such mentions, 35% are attributable to our sample leaders. Panels (c) and (d) compare thematic emphases and tone between in-sample and out-of-sample politicians: (i) about 55% of interventions discuss "community cohesion" (against 15% for "security"); (ii) these topics do fluctuate over time, as does the positivity of the mention, only in part driven by major conflicts; (iii) in-sample leaders tend to discuss "community cohesion" more often and adopt more moderate and positive rhetoric.

Figure A12. Speeches from the Knesset and the "upbringing" of speakers.



(a) Cohesion (b) Tone (c) Arab-Jewish relations

Notes: These figures display the correlation between our shift-share uniformity prediction, $\widetilde{\vartheta}$, associated with leaders of our sample and the nature of Knesset interventions. In panel (a), we focus on the likelihood to intervene on community cohesion and the absorption of immigration waves. In panel (b), we consider the likelihood to intervene in a constructive or optimistic tone. In panel (c), we only consider mentions of "Arab-Jewish relationships" and report the average positivity score (on a scale from 0 to 100) associated with these mentions.

Speeches and the upbringing of speakers In this final section, we explore the relationship between the speakers' backgrounds and the content and tone of their interventions. In the spirit of our exercises of Section 4.2, we contrast features characterizing the nature of Knesset interventions with our shift-share measure of settlement uniformity, $\widetilde{\vartheta}$, in the settlements where the speakers were raised/hosted. Figure A12 (a) shows that individuals from more diverse settlements are much more likely to emphasize themes of community cohesion and immigration absorption: the incidence of such topics declines from approximately 0.25 in the most diverse settlements to 0.10 in the least. Figure A12 (b) indicates that leaders from diverse communities are also more likely to adopt a constructive tone—60% of their interventions, compared to 40% for leaders from less diverse backgrounds. This pattern is echoed in Figure A12 (c), which focuses exclusively

on mentions of Jewish-Arab relations: leaders from diverse settlements exhibit markedly higher positivity scores than their counterparts from more homogeneous communities.

Table A1. Geography and suitability to host a settlement.

Settlement density	(1)	(2)
Hexagon longitude (ITM)	-0.00754	-0.03007
	(0.04744)	(0.01214)
Hexagon latitude (ITM)	-0.01038	0.05152
	(0.02846)	(0.00729)
Elevation (mean)	-0.00221	-0.00936
	(0.00906)	(0.00232)
Slope	-0.00439	-0.04429
	(0.10694)	(0.02737)
Soil bulk density	0.03088	0.04412
	(0.04356)	(0.01115)
Soil organic carbon content	-0.53166	0.58060
	(1.04463)	(0.26738)
Average precipitation	0.08359	0.08131
	(0.15794)	(0.04043)
Average temperature	-0.00687	0.00011
	(0.15356)	(0.03931)
Area occupied by water (sq. km)	0.11182	0.35312
	(0.67861)	(0.17369)
Inverse distance to the nearest shore	-4.10956	-2.89277
	(4.88098)	(1.24931)
Inverse distance to the nearest canal	-0.73343	-6.22068
	(3.38260)	(0.86579)
Inverse distance to the nearest water	4.97365	0.04478
	(1.60920)	(0.41188)
Yield of wheat (rain)	0.00010	-0.00110
	(0.00106)	(0.00027)
Yield of sorghum (rain)	0.00035	-0.00015
	(0.00126)	(0.00032)
Yield of maize (rain)	-0.00006	0.00064
	(0.00091)	(0.00023)
Yield of citrus (rain)	-0.00062	-0.00039
	(0.00116)	(0.00030)
Yield of vegetables (rain)	-0.00022	0.00146
	(0.00112)	(0.00029)
Yield of cotton (rain)	-0.00169	-0.00173
	(0.00815)	(0.00209)
Inverse travel cost to Haifa	1.74176	72.11554
	(10.45540)	(2.67611)
Inverse travel cost to Jerusalem	1.66769	42.73700
	(12.12478)	(3.10339)
Inverse travel cost to Tel Aviv	35.12910	49.41977
	(14.75657)	(3.77701)
Inverse travel cost to any Arab settlement	8.14399	2.08593
	(9.60424)	(2.45825)
Inverse distance to the nearest railways	-6.77366	3.42359
	(3.26947)	(0.83683)
Inverse distance to the nearest roads	-0.02669	1.37240
	(1.38487)	(0.35446)
Share of land claimed by the Committee for Palestine	14.61885	-0.63331
	(4.49620)	(1.15082)
Share of land claimed by the Jewish National Fund	77.59722	3.13662
	(4.71518)	(1.20687)
Share of land regulated as free	12.83527	0.54950
	(5.25750)	(1.34568)
Share of land regulated as zone A	-0.50120	1.01356
	(3.16855)	(0.81101)
Share of land regulated as zone B	0.60339	1.81770
	(3.41771)	(0.87478)
Observations	25,040	25,040
R-squared	0.02923	0.53993

Notes: A unit of observation is a hexagon with a radius of 2 kilometers. Robust standard errors are reported between parentheses.

A.5 Predicting settlement creation

This section provides complements to Section 3.1 predicting the dynamic creation of settlements in Mandatory Palestine.

Suitability for settlement In Section 3.1, we discuss how we construct an index of settlement suitability, $\widehat{p}_i$, based on geographic characteristics within hexagon i. We provide an illustration of its geographic variation across Mandatory Palestine in Figure 2. We report the detailed prediction in Table A1. In column (1), the dependent variable is settlement creation within a hexagon over the period 1920–1947. In column (2), the dependent variable is the change in proximity to settlements between 1920–1947 ($s_{i,47} - s_{i,20}$): our suitability variables, e.g., proximity to large cities, to water, to transportation infrastructure, agricultural yields, or topography, explain about 54% of the variance in the geography of settlement creation over the period (column 2).

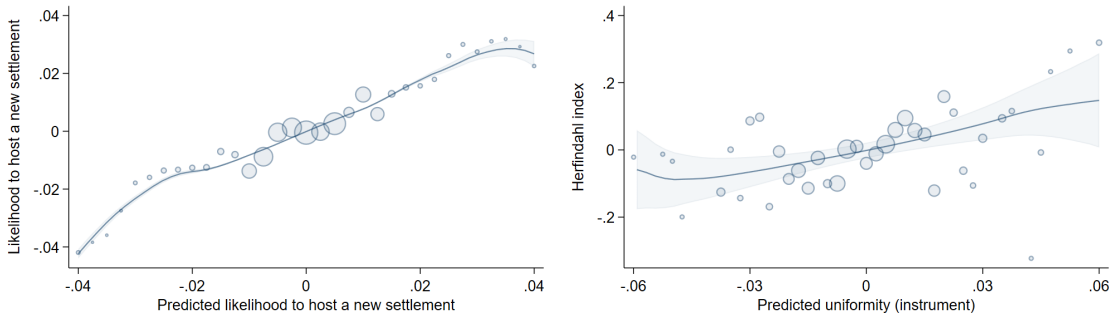
Table A2. Predicting the timing of settlement creation—robustness to forward values.

Change in settlement density ($\Delta s_{i\tau}$)	(1)	(2)	(3)
Predicted change in settlement density	1.314	0.884	0.707
	(0.102)	(0.071)	(0.051)
Predicted change in settlement density (forward)	-0.433	0.424	0.313
	(0.107)	(0.087)	(0.069)
Observations	22,536	22,536	22,536

Notes: A unit of observation is a hexagon with a radius of 2 kilometers in a given time period (every three years between 1920–1947). All specifications include hexagon fixed effects and time fixed effects, and are specified in differences (for the dependent variables and the main explaining variable)—see Equation (3). In column (2), we control for hexagon-specific time trends. In column (3), the explaining variable is the normalized variable $\widetilde{s}_{i\tau}$ (and its forward value)—thus providing a direct validation for the construction of the "shares" in the shift-share prediction (4). Standard errors, computed following Müller and Watson (2022) to account for spatial auto-correlation, are reported between parentheses.

Predicting the timing of settlement creation In Table 1, we report the estimation of Equation (3) and show that changes in *actual* settlement density correlate with changes in *predicted* settlement density, even conditioning on hexagon-fixed effects. We shed further light on the timing of settlement creation in Table A2, where we regress changes in *actual* settlement density on changes in *predicted* settlement density and its forward value. Reassuringly, we find the contemporaneous prediction to be more predictive than the forward prediction.

Figure A13. Illustration of the first-stage relationship(s)—predicting the timing of settlement creation and uniformity within the settlement.



(a) Timing of settlement creation

(b) Predicting uniformity

Notes: These figures provide an illustration of the relationships displayed in Table 1 (column 3) and Table 2 (column 2). In panel (a), we report the relationship between the normalized variable $\widetilde{s}_{i\tau}$ and the actual increase in local settlement density, $\Delta s_{i\tau}$—the adjusted R-squared associated with column (3) of Table 1 is 0.88. A unit of observation is a hexagon with a radius of 2 kilometers in a given time period (every three years between 1920–1947), and both variables are residualized by hexagon fixed effects, time fixed effects, and hexagon-specific trends. In panel (b), we report the relationship between the uniformity measure constructed from (i) the timing of settlement creation as predicted by the "frontier expansion" algorithm and (ii) a Herfindahl index of yearly migrant flows across countries of origin and the Herfindahl index of origins across all settlers as measured in 1945—the adjusted R-squared associated with column (2) of Table 2 is 0.32. Both variables are residualized by all separate geographic characteristics used to capture suitability, the initial allocation of settlements and travel time to settlements around 1920, a trend and quadratic term in the year of establishment, population-decile fixed effects, and fixed effects at the level of sub-districts. For the sake of visualization, we group observations by bins and display a local polynomial fit.

Illustration of the first-stage relationship(s) The main idea behind our empirical strategy is to combine a predictor for settlement creation, denoted $\widetilde{s}_{i\tau}$, with the period-specific composition of migration inflows, θ_τ, into a shift-share predictor, $\widetilde{\vartheta}_i = \sum_\tau \widetilde{s}_{i\tau} \times \theta_\tau$. In Figure A13, we provide a visualization of two key validation relationships: (a) the relationship between the prediction, $\widetilde{s}_{i\tau}$, and the actual increase in local settlement density, $\Delta s_{i\tau}$ (panel a—associated with the estimate presented in column 3 of Table 1); and (b) the relationship between predicted uniformity ($\widetilde{\vartheta}_i$) and actual measures of uniformity within settlement i (panel b—associated with the estimate presented in column 2 of Table 2). The very tight and precise relationship between predicted and actual settlement creation (Figure A13, panel a) is what allows us to keep significant explanatory power when combining such predicted settlement creation with aggregate immigrant characteristics (Figure A13, panel b).

Shedding light on the dynamics of settlement creation In Section 3.1, we discuss how we predict the dynamics of settlement creation using an iterative process (see Equation 2). We report the empirical estimates of Equation (2) in Table A3 where we see that settlement creation follows an inverted U-pattern in previous settlement density (with a positive coefficient in front of *Settlement density*, but a negative one in front of squared density). In other words, settlements appear at an intermediate distance from previous

Table A3. Dynamics of settlement creation.

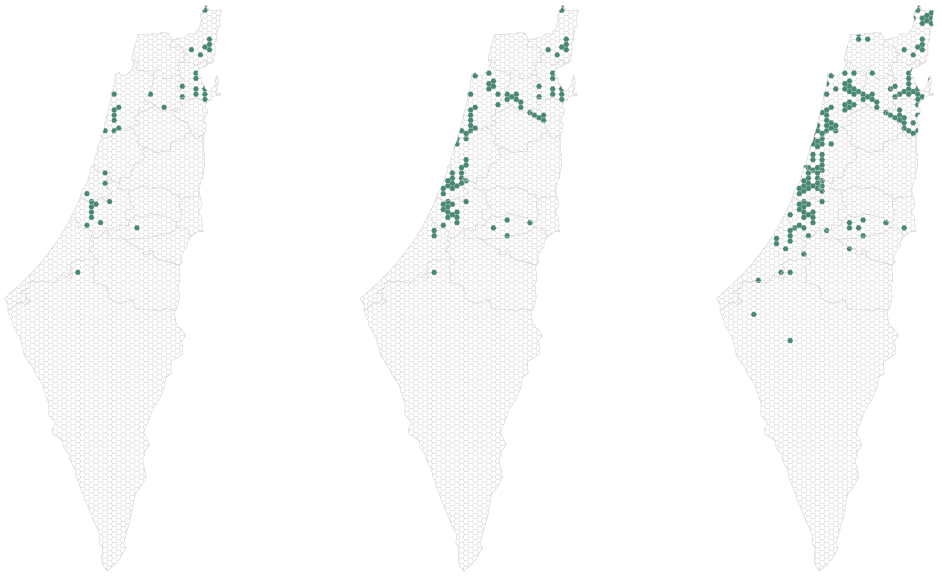
Settlement density	(1)	(2)
Suitability	0.044	0.056
	(0.048)	(0.045)
Settlement density ($t-1$)	1.075	1.064
	(0.017)	(0.017)
Suitability × Settlement density ($t-1$)	4.929	4.933
	(0.248)	(0.266)
Settlement density (sq., $t-1$)	-0.332	-0.277
	(0.048)	(0.046)
Suitability × Settlement density (sq. $t-1$)	-2.287	-2.859
	(0.441)	(0.444)
Observations	25,040	26,700
R-squared	0.993	0.992

Notes: A unit of observation is a hexagon with a radius of 2 kilometers. Suitability is constructed using: latitude and longitude; maximum, minimum and average elevation; average slope; soil bulk density at 0 cm depth as reconstructed from recent satellite imagery (Hengl, 2018); organic carbon content (Hengl and Wheeler, 2018); average annual precipitation and temperature (Hijmans et al., 2005); distance to the nearest shore, river, canal, stream, wadi; the length and area of water bodies; the distance to and length of roads and railways as of 1920; the potential yield per hectare for various crops (citrus in particular) as extracted from the Global Agro-Ecological Zones Agricultural Suitability and Potential Yields (GAEZ); the distance and travel time to the nearest existing city and the distance and travel time to Haifa, Jerusalem and Tel-Aviv/Jaffa; the density, distance/travel time to existing Arab settlements; the share of land claimed by the Jewish companies and private owners and by the Jewish National Fund; and later land regulations imposed by the British mandate. See Section 3.1, or Appendix Table A1. In column (1), suitability is based on the full list of controls. Suitability is constructed from a parsimonious set of controls in column (2): latitude, longitude, average elevation, organic carbon content, distance to the nearest shore, canal, yield per hectare for citrus and cotton, the distance to roads and railways as of 1920, and land owned by the Jewish National Fund. Standard errors, computed following Müller and Watson (2022) to account for spatial auto-correlation, are reported between parentheses.

settlements. Settlements also appear in locations that are deemed suitable, especially those that are in the proximity of established settlements—*Suitability* × *Settlement density*—but not in their immediate proximity—*Suitability* × *Settlement density (sq.* $t-1$*)*. This non-linearity and its dependence on suitability is best shown in Figure 3 of the paper, illustrating the nature of the mapping $f : (\mathbf{s}) \mapsto f(\mathbf{s})$ as a function of different suitability levels. In column (1) of Table A3, suitability is based on the full list of controls; and suitability is constructed from a parsimonious set of controls in column (2).

We further illustrate the dynamics of actual settlement creation and predicted settlement creation in Figures A14 and A15. Figure A14 shows the initial allocation of settlements and the *actual* dynamics of settlement creation. In 1920, the settlements do not form a continuous frontier, but are grouped in a few clusters around key, strategic locations. The expansion of the frontier between 1920 and 1932 is straightforward:

Figure A14. Settlements in Mandate Palestine.



(a) 1920 (b) 1932 (c) 1944

Notes: Panels (a), (b) and (c) show the distribution of settlements in 1920, in 1932 and in 1944 across Mandate Palestine. The main geographic unit is a hexagon with a radius of 2 kilometers; green hexagons indicate that there is at least the centroid of one settlement (either agricultural or mainly owned by the Histadrut—the General Organization of Workers in Israel) within the hexagon.

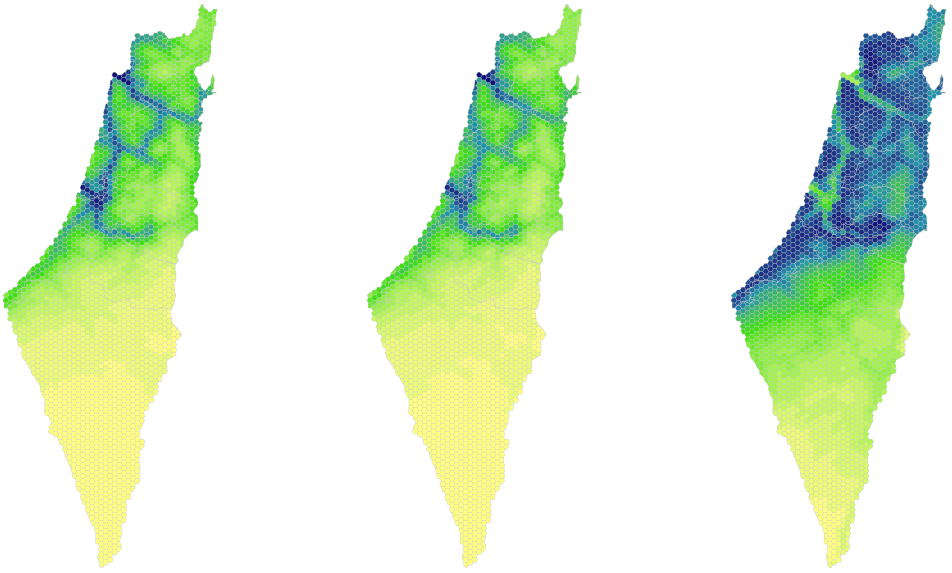
settlements appear along the main trading routes in between the existing settlements—starting to form a continuous frontier along the coast, extending into the hinterlands from Haifa towards Nazareth, and then from Nazareth to the North of the country. Between 1932 and 1944, however, new settlements cannot be located just along trading routes and they need to stray away while remaining at a reasonable distance. The objective of our "frontier expansion" algorithm is to exploit this dynamics of settlement creation, but to make it entirely dependent on the initial allocation (and thus ignore idiosyncratic, and possibly endogenous settlement creation between 1920–1947).

Figure A15 displays the heatmap of settlement creation as predicted by the "frontier expansion" algorithm in 1920, 1932 and 1944. One can see that the frontier is expected to expand along trading routes in 1932 because there remain many "holes" in very suitable locations. However, this pattern changes markedly in 1944 because there are already too many settlements along the main axes: the new settlements are predicted to appear in the proximity of this continuous frontier, but away from the main trading routes. In summary, our "frontier expansion" algorithm does capture the major patterns of actual settlement creation over time, as shown in Figure A14:

- the importance of (given) geographic factors;

- and the dynamics of settlement creation, as a function of previous settlement geographies.

Figure A15. Predicting settlement creation in Mandate Palestine.



(a) 1920 (b) 1932 (c) 1944

Notes: The main geographic unit is a hexagon with a radius of 2 kilometers. Panels (a), (b) and (c) show the distribution of marginal settlement density by 1920, 1932, and 1944 across Mandate Palestine. The pink dots represent actual settlement creation during the period. Colors indicate the predicted marginal probability to have a settlement in the previous three years (ranging from 0 in light yellow to higher values in dark blue while green indicates intermediate values). The procedure to create these predictions is described in Section 3.1.

B Complements to the empirical analysis

This section provides complements to the empirical analysis, mostly expanding on a summary provided in Section 4.1. More specifically,

- we provide supporting evidence about the first stage of our strategy (see Section 3.2 of the paper) in Appendix B.1;
- we conduct a sensitivity analysis of our baseline specification in Appendix B.2;
- we provide support for the identification hypothesis in Appendix B.3;
- we discuss the selection of leaders, their allocation across settlements and provide complementary evidence on their rise in Appendix B.4;
- we consider alternative outcomes for our study of inter-ethnic relationships (as a complement to Section 4.2) in Appendix B.5;
- settlement heterogeneity and sample selection in Appendix B.6.
- we discuss community and nation building within the kibbutz (as a complement to Section 5.1) in Appendix B.7.

B.1 The shift-share prediction and the first-stage

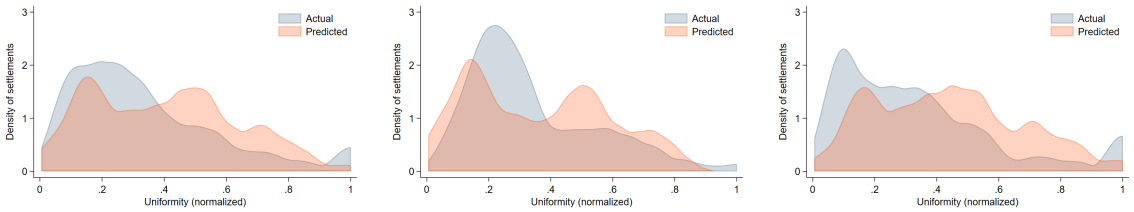
Our empirical design uses a shift-share prediction to allocate migrants into settlements. In this section, we provide additional empirical support for this strategy.

Validating the shift-share prediction of uniformity within settlements As described in Section 3.1, we predict uniformity at the settlement level by leveraging the timing of settlement creation interacted with the composition of migrant flows. Letting θ_τ denote the Herfindahl index across origins of all immigrants arriving between $\tau - 1$ and τ, we translate this uniformity-index of the average *immigrant* into a uniformity-index of the average *settler* in just-created settlements (ϑ_j, within settlement j).[42]

In Figure B1, we report the distribution of uniformity measures used in the main analysis of the paper (see, e.g., Table 2) for all settlements (panel a), kibbutzim (panel b), and the other settlement organizations (panel c). There are two main messages conveyed by this figure. First, our prediction has a fatter tail than the actual distribution of Herfindahl indices across origins, meaning that we slightly under-estimate the extent of diversity

[42] Concretely, we use a probabilistic measure to perform this operation such that the term "just-created settlements" is not correct: we assign immigrants on the basis of the period-specific probabilities that a location receives a new settlement, possibly before or after their actual creation (if any).

Figure B1. Distribution of predicted and actual uniformity measures.

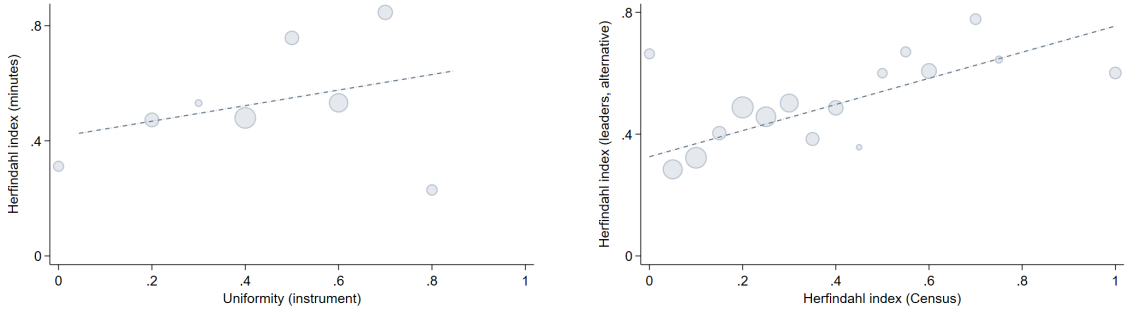


(a) All settlements (b) Kibbutzim (c) Others

Notes: This Figure displays the distribution of uniformity measures used in the main analysis of the paper (see, e.g., Table 2) for all settlements (panel a), kibbutzim (panel b), and the others (panel c). The *actual* Herfindahl index of origins is calculated across all settlers present in 1945; the *predicted* uniformity is a shift-share prediction based on the timing of settlement creation as predicted by the "frontier expansion" algorithm and a Herfindahl index of yearly migrant flows.

within settlements. Second, kibbutzim tend to be more diverse than other settlement types, a feature that our instrument is not picking up: our instrument is purely based on the "frontier expansion" coupled with aggregate inflows to Palestine, thus missing the distinct appeal of kibbutzim to settlers of different characteristics. In short, kibbutzim might be attractive to settlers for other reasons than the origins of established members (because of their structure and/or because of the preferences of their targeted group), and this appeal might differ from the one of other settlement types. By construction, our instrument however ignores any such (endogenous) differences across settlements.

Figure B2. Validation of uniformity measures.



(a) Predicted uniformity (b) Actual uniformity

Notes: Panel (a) compares a Herfindahl index of migrant flows as reconstructed from the minutes of meetings to our shift-share instrument based on the timing of settlement creation as predicted by the "frontier expansion" algorithm and a Herfindahl index of yearly migrant flows. Panel (b) compares a Herfindahl index of origins based on influential settlers (6,000 from the *Encyclopedia of the Founders and Builders of Israel*) to the actual Herfindahl index of origins, as calculated across all settlers present in 1945. For exposition purposes, we group observations in dots by deciles of the x-axis variables, weighted by the number of mentions, and we show the best linear fit(s).

In Figure B2, we provide indirect, supporting evidence that our instrument does allocate migrants in a meaningful manner. In panel (a), we correlate our measure with a

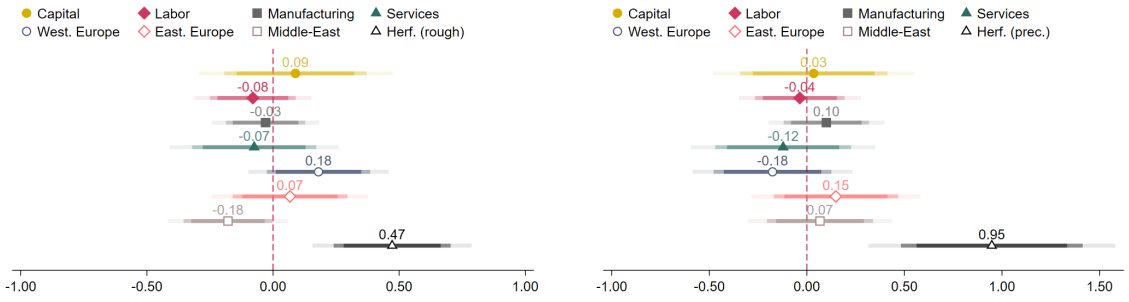
Herfindahl index of migrant flows as reconstructed from the minutes of meetings. Indeed, in our minutes of meetings described in Section 2, there are explicit references to groups arriving in different settlements and their origins. We provide examples here (destination; origin): "There was another group of immigrants from Hungary, Transylvania, who went to Kfar Sold" (Kfar Szold; Hungary, Transylvania); "A group from Kielce came to us, a selected group. They went with a heavy heart to Herzliya. They asked to give them a chance to succeed elsewhere" (Herzliya; Poland); "[David] now offers 8 families with 11 children. We are interested in taking them in order to be a force of attraction for German immigration" (Kinneret; Germany). These discussions show that the allocation is disciplined by the preferences of immigrants, but also considerations from receiving communities. Our instrument ignores these dimensions, and rather constructs an allocation mechanism purely based on "vacancies", i.e., newly-created settlements are in need of labor and are ready to absorb people. We systematically identify such references and extract the settlement of destination, the number of migrants when specified and their origin. From these fragmented, scattered mentions, we can construct a Herfindahl index of migrant flows and compare it with our shift-share prediction. Naturally, those references are far from being exhaustive, and the measure is probably noisy (see panel a of Figure B2).

In panel (b) of Figure B2, we construct a Herfindahl index of origins based on influential settlers (6,000 from the *Encyclopedia of the Founders and Builders of Israel*) rather than on the census of all settlers (as we do in our baseline strategy). This check serves two purposes: (i) a validation of the census data, diversity in the census does seem to be correlated with diversity as extracted from settlers' biographies, and (ii) an illustration that influential settlers do select in a correlated fashion with other settlers.

The uniformity of successive migrant waves Successive migration waves differing in their diversity/uniformity may also differ in important factors of settlement in Palestine: the wealth of settlers, their regions of provenance, or their skills and occupations. We discuss how these confounders might, in principle, affect the *interpretation* of our main estimate(s) of interest in Section 3.2.

Our preferred strategy to quantify their importance will be to control for these confounders in an extended version of our baseline specification. In this section, we shed *some* light on the relevance of these confounders by correlating the Herfindahl index of migrant inflows over the years between 1920 and 1947 and other characteristics of these migrant waves: the wealth of settlers, their previous occupations, and their origins (Western Europe, Central/Eastern Europe, Middle-East). Figure B3 shows that these correlations remain limited.

Figure B3. The uniformity of migrant waves over time.



(a) Detailed Herfindahl (countries)

(b) Rough Herfindahl (regions)

Notes: This Figure measures the correlations between Herfindahl indices of migrant inflows over time (computed across all countries of origin in panel a, and across 8 main regions of origin in panel b) and other characteristics of these migrant waves: the share of immigrants admitted as capitalists; the share of immigrants admitted as laborers; the share of migrants with previous occupations in manufacturing; the share of migrants with previous occupations in services; the share of migrants from Western Europe, including Germany; the share of migrants from Central/Eastern Europe; and the share of migrants from the Middle-East. The reported estimates are those of a regression where each "share" is regressed against the Herfindahl index of migrant inflows over time, controlling for a second-order polynomial in the year (as indirectly done in the baseline specification through the time of settlement creation). The last row shows the estimate(s) between the two Herfindahl indices of migrant inflows.

Table B1. Predicting settlement composition across settlements—sensitivity analysis.

Herfindahl index (ϑ_i)	(1)	(2)	(3)
Predicted uniformity ($\widetilde{\vartheta}_i$)	1.261	1.426	1.159
	(0.251)	(0.267)	(0.276)
Observations	316	316	316
Construction	Rough origins	Detailed origins	Herfindahl

Notes: A unit of observation is a settlement. The dependent variable is a Herfindahl index of origins across all settlers; the explaining variables are uniformity measures based on the timing of settlement creation as predicted by the "frontier expansion" algorithm and yearly migrant flows. All specifications include the variables used in the "frontier expansion" algorithm (see Section 3.1): all separate geographic characteristics used to capture suitability; and the initial allocation of settlements and travel time to settlements around 1920. We also include fixed effects at the level of sub-districts, a trend and quadratic term in the year of establishment, as well as population-decile fixed effects. In column (1), we construct the shift-share prediction from 8 large regions of origin—the same as those used by census enumerators and used to construct the Herfindahl index of origins across all settlers. In column (2), we construct the shift-share prediction from 14 more detailed regions of origin. In column (3), the measure of uniformity is a non-linear construction from (linearly) predicted shares of each migrant group. Standard errors are clustered at the sub-district level and reported between parentheses.

First-stage specification We now provide a sensitivity analysis of our first-stage specification. In Table B1, we consider the specification shown in column (2) of Table 2 but replace the baseline shift-share uniformity measure by: (i) a shift-share prediction based on the 8 large regions of origin used by census enumerators and used to construct the Herfindahl index of origins across all settlers; (ii) a shift-share prediction based on 14 regions of origin; and (iii) a non-linear Herfindahl index built from (linearly) predicted shares of each migrant group. Table B1 shows that these modifications do not weaken

the first stage too significantly; the next section will discuss how they affect our main estimates.

Table B2. Predicting settlement composition across settlements—robustness to controls.

Herfindahl index (ϑ_i)	(1)	(2)	(3)	(4)	(5)
Predicted uniformity ($\widetilde{\vartheta}_i$)	2.226	2.445	2.612	1.726	2.235
	(0.506)	(0.295)	(1.498)	(0.301)	(0.325)
Observations	301	316	183	316	316
Controls	Composition	Origins	Capital	Type	Market

Notes: A unit of observation is a settlement. The dependent variable is a Herfindahl index of origins across all settlers; the explaining variable is a uniformity measure based on the timing of settlement creation as predicted by the "frontier expansion" algorithm and a Herfindahl index of yearly migrant flows. All specifications include the variables used in the "frontier expansion" algorithm (see Section 3.1): all separate geographic characteristics used to capture suitability; and the initial allocation of settlements and travel time to settlements around 1920. We also include fixed effects at the level of sub-districts, a trend and quadratic term in the year of establishment, as well as population-decile fixed effects. In column (1), we control for the shares of: settlers speaking a foreign language as first language, literate settlers, kibbutz workers, white collars, unskilled workers, and capitalists. In column (2), we control for the separate shares of settlers from Eastern Europe, the Balkans, Western Europe, Asia, Africa, America/Oceania. In column (3), we control for the capital stock (and indebtedness) in 1945 at the settlement level and the share of local land owned by Jews (versus the Arab population). In column (4), we control for 4 settlement types (Kibbutz, Moshav, City, Other) and 8 umbrella organizations (e.g., Kibbutz Artzi). In column (5), we control for a measure of market access to other Jewish settlements in 1947. Standard errors are clustered at the sub-district level and reported between parentheses.

We further probe the robustness of our first-stage relationship in Table B2 where we control for other dimensions of settlement composition and finer geographical characteristics. Table B2 shows that our first-stage estimates are not affected by the inclusion of controls for precise measures of settlement composition (language, education, occupation, origins), capital, settlement types and measures of market access. While this evidence provides some empirical, "practical" support for the orthogonality of our instrument with respect to some of these confounding factors, our main "theoretical" argument to support exogeneity is our "frontier expansion" algorithm, which replaces the actual formation of a community in a certain location by a dynamic probability heatmap over Mandatory Palestine. The dynamic process is informed by natural amenities and by a recursive process initiated by the location of historical Jewish communities.

B.2 Sensitivity analysis around the baseline specification

This section provides a sensitivity analysis of our baseline specification and discusses robustness to: variations around the baseline specification (e.g., other thresholding procedures for dependent variables, different inference accounting for spatial correlation); and the addition of important confounding factors (education, occupation and origin of the average settler; settlement type; second-nature geography).

Table B3. Diversity, the rise of societal leaders and inter-ethnic relationships—OLS specification.

	All leaders (1)	Mil. leaders (2)	Pol. leaders (3)	Cul. leaders (4)	Casualties (5)	Pacts (6)
Herfindahl	-0.621	-0.165	-0.407	-0.047	0.026	-0.273
	(0.178)	(0.073)	(0.107)	(0.026)	(0.057)	(0.078)
Observations	316	316	316	316	316	316

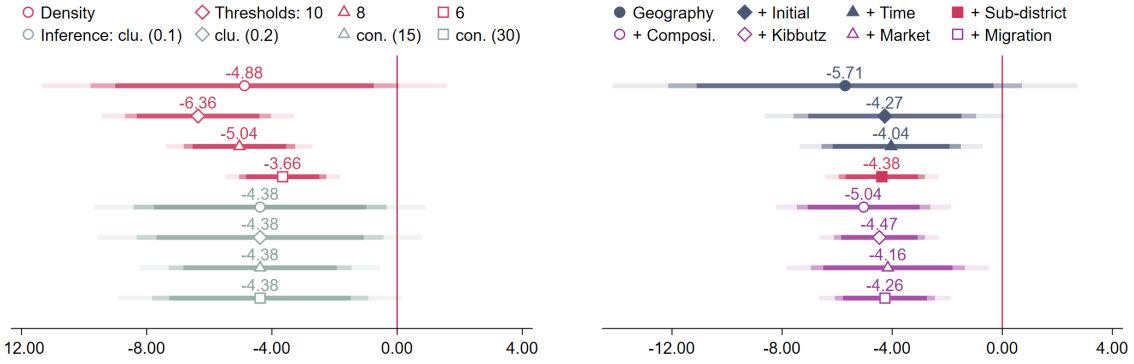
Notes: A unit of observation is a settlement. The dependent variable is: the number of military, political, and cultural leaders having resided in the settlement in column (1); the number of military leaders in column (2); the number of political leaders in column (3); the number of cultural leaders in column (4); the probability of civilian casualties in nearby Arab villages within 10 kms in column (5); and the total number of pacts between each Jewish settlement of our sample and neighboring Arab villages in column (6). The explaining variable is a Herfindahl index of origins across all settlers. All specifications include the control variables used in Table 3; we also control for the Arab population and the presence of military Arab troops in these nearby villages. Standard errors are clustered at the sub-district level and reported between parentheses.

OLS specification In Table B3, we report OLS estimates which confirm the direction of our main effects, but their magnitudes sometimes differ from the IV specification. This pattern holds across our primary outcomes: the emergence of societal leaders, the incidence of civilian casualties, and the number of pacts formed with Arab villages. A likely explanation is classical measurement error—for example, our diversity measure is captured at a single point in time (1945)—and the resulting attenuation bias observed in OLS specifications. Alternative explanations include potential reverse causality: influential individuals may attract others from similar backgrounds, thereby reducing observed diversity (for the estimates on leadership); and diverse settlements may be more likely to arise in areas with pre-existing tensions with the Arab population, potentially confounding conflict-related estimates.

Thresholding, inference, and baseline controls Our main "leader" estimate (column 1 of Table 3) is based on: a transformation of the main outcome variables, e.g., the (truncated) number of leaders having resided in a settlement; a predicted Herfindahl index based on a wide set of predictors for suitability; and clustering at the level of sub-districts. In Panel (a) of Figure B4, we show that our results are not changed by normalizing the number of leaders by population (first row), and by a few settlements with many leaders (rows 2-4 using different truncation thresholds). We also consider different approaches to inference [green]: clustering at the level of a 0.1 degree × 0.1 degree latitude/longitude cell, 0.2 degree × 0.2 degree latitude/longitude cell, and Conley standard errors (15 kms, 30 kms).

The "shares" of our shift-share design partly relies on first- and second-nature geography combined with the location of initial settlements. In Figure B4 (panel b), we discuss the importance of controlling for these separate components as well as for smooth func-

Figure B4. Diversity and the rise of societal leaders—thresholding and robustness to controls.



(a) Thresholding/inference

(b) Controls

Notes: This Figure reports a sensitivity analysis for the "leader" estimate (column 1 of Table 3). In panel (a), we use a different normalization and different thresholds for our leader variable [red], and we consider different approaches to inference [green], most notably clustering at the level of a 0.1 degree × 0.1 degree latitude/longitude cell, 0.2 degree × 0.2 degree latitude/longitude cell, and Conley standard errors (15 kms, 30 kms). In panel (b), we first consider specifications with fewer controls than in the baseline [blue]: geographic variables used for suitability only, adding initial conditions (settlement density and initial population), adding year of establishment (trend and quadratic terms), adding sub-district fixed-effects—the estimate then coincides with column (1) of Table 3 [red]. We also consider specifications with more controls than in the baseline [purple], adding separately: shares per origin in 1945, settlement-type fixed effects interacted with year of establishment (trend and quadratic terms), measures of market access computed in 1947, and predicted migration *levels*.

tions in the year of settlement creation, and controlling for geographic differences that are not well accounted for in the baseline regression. Intuitively, we slowly converge to our baseline regression (in red) by successively adding the location of initial settlements (row 2), quadratic functions of the timing of settlement creation (row 3), and sub-district fixed effects (row 4—the estimate then coincides with column 1 of Table 3). In a second step, we consider specifications with more controls than in the baseline [purple in Figure B4], adding separately: shares per origin in 1945, year of establishment (trend and quadratic terms), measures of market access computed in 1947, and predicted migration *levels*. Our results are robust to these variations, showing that they are not dependent on certain origins, a better (ex-post) "geography" of diverse settlements, or a migration incidence effect whereby diverse settlements would be those with higher immigration rate. More specifically, we show that the main "leader" estimate (column 1 of Table 3) remains quite stable across these different tests, noting that the other estimates are equally robust [unreported checks].

Specification of the shift-share prediction of uniformity In Table B4, we report a sensitivity analysis to the choice of instrument and use: a uniformity measure based on a parsimonious set of geographic variables—as in Table A3—in column (1); a uniformity measure based on polarization indices of the successive migration waves in column (2); a

Table B4. Diversity and the rise of societal leaders—alternative shift-share predictions of uniformity.

Instrument is ...	All leaders (1)	Mil. leaders (2)	Pol. leaders (3)	Cul. leaders (4)	Casualties (5)	Pacts (6)
Parsimonious	-2.626	-0.896	-0.691	-1.038	1.455	-0.605
	(0.885)	(0.375)	(0.346)	(0.314)	(0.496)	(0.435)
	[51.37]	[51.37]	[51.37]	[51.37]	[51.37]	[51.37]
Polarization	-3.638	-1.244	-1.125	-1.269	1.419	-0.760
	(0.983)	(0.431)	(0.282)	(0.361)	(0.450)	(0.436)
	[70.58]	[70.58]	[70.58]	[70.58]	[70.58]	[70.58]
Regions of origin (8)	-5.100	-1.852	-1.472	-1.776	1.684	-0.937
	(1.487)	(0.559)	(0.422)	(0.624)	(0.578)	(0.483)
	[30.64]	[30.64]	[30.64]	[30.64]	[30.64]	[30.64]
Regions of origin (14)	-4.908	-1.771	-1.433	-1.704	1.654	-0.909
	(1.404)	(0.538)	(0.397)	(0.583)	(0.555)	(0.474)
	[36.08]	[36.08]	[36.08]	[36.08]	[36.08]	[36.08]
Herfindahl	-5.856	-2.168	-1.665	-2.023	1.842	-1.028
	(1.806)	(0.656)	(0.528)	(0.769)	(0.657)	(0.516)
	[19.86]	[19.86]	[19.86]	[19.86]	[19.86]	[19.86]

Notes: Each row reports the outcome of a separate regression in which a unit of observation is a settlement, and the dependent variable is: the number of leaders having resided in the settlement in column (1); military leaders in column (2); political leaders in column (3); cultural leaders in column (4); the probability of civilian casualties in nearby Arab villages within 10 kms in column (5); and the total number of pacts between each Jewish settlement of our sample and neighboring Arab villages in column (6). The explaining variable is a Herfindahl index of origins across all settlers; the instrument is: a parsimonious uniformity measure in the first row with suitability inferred from latitude, longitude, average elevation, organic carbon content, distance to the nearest shore, canal, yield per hectare for citrus and cotton, the distance to roads and railways as of 1920, and land owned by the Jewish National Fund; a uniformity measure based on polarization indices of the successive migration waves in the second row; a shift-share prediction from 8 large regions of origin—the same as those used by census enumerators and used to construct the Herfindahl index of origins across all settlers—in the third row; a shift-share prediction from 14 more detailed regions of origin in the fourth row; and a non-linear construction from (linearly) predicted shares of each migrant group in the last row. All specifications include the variables used in the "frontier expansion" algorithm (see Section 3.1): all separate geographic characteristics used to capture suitability; and the initial allocation of settlements and travel time to settlements around 1920. We also include a trend and quadratic term in the year of establishment, as well as population-decile fixed effects and fixed effects at the level of sub-districts. Standard errors are clustered at the sub-district level and reported between parentheses. F-statistics are reported between brackets.

shift-share prediction from the same 8 large regions of origin used in the actual Herfindahl index across settlers; a shift-share prediction from 14 more detailed regions of origin; and a non-linear construction from (linearly) predicted shares of each migrant group. All these objects are based on the timing of settlement creation as predicted by the "frontier expansion" algorithm and the composition of successive migrant flows. The estimates are not too dissimilar from our baseline estimates shown in Tables 3 and 4.

In the next subsection, we probe the robustness of our results to more demanding tests, focused on identification concerns.

Table B5. Diversity and the rise of societal leaders—robustness to controlling for predicted and actual settlement composition.

Controlling for ...	All leaders (1)	Mil. leaders (2)	Pol. leaders (3)	Cul. leaders (4)	Casualties (5)	Pacts (6)
Panel A: Predicted capitalists						
	-5.454	-1.990	-1.608	-1.856	1.792	-0.951
	(1.478)	(0.587)	(0.422)	(0.612)	(0.586)	(0..494)
	[24.33]	[24.33]	[24.33]	[24.33]	[24.33]	[24.33]
Panel B: Predicted manufacturing						
	-2.012	-0.573	-0.712	-0.727	1.100	-0.577
	(0.521)	(0.279)	(0.252)	(0.096)	(0.341)	(0.422)
	[66.38]	[66.38]	[66.38]	[66.38]	[66.38]	[66.38]
Panel C: Predicted migration						
	-3.023	-0.986	-0.992	-1.044	1.319	-0.678
	(0.637)	(0.328)	(0.239)	(0.169)	(0.394)	(0.418)
	[89.93]	[89.93]	[89.93]	[89.93]	[89.93]	[89.93]
Panel D: Occupational structure and literacy						
	-3.329	-1.127	-0.872	-1.328	1.444	-0.799
	(0.993)	(0.458)	(0.329)	(0.396)	(0.489)	(0.493)
	[46.88]	[46.88]	[46.88]	[46.88]	[46.88]	[46.88]
Panel E: Group dynamics						
	-4.076	-1.393	-1.241	-1.441	1.763	-0.856
	(1.027)	(0.479)	(0.281)	(0.394)	(0.554)	(0.546)
	[28.47]	[28.47]	[28.47]	[28.47]	[28.47]	[28.47]

Notes: Each row reports the outcome of a separate regression in which a unit of observation is a settlement, and the dependent variable is: the number of leaders having resided in the settlement in column (1); military leaders in column (2); political leaders in column (3); cultural leaders in column (4); the probability of civilian casualties in nearby Arab villages within 10 kms in column (5); and the total number of pacts between each Jewish settlement of our sample and neighboring Arab villages in column (6). The explaining variable is a Herfindahl index of origins across all settlers; the instrument is our shift-share prediction. All specifications include the variables used in the "frontier expansion" algorithm (see Section 3.1): all separate geographic characteristics used to capture suitability; and the initial allocation of settlements and travel time to settlements around 1920. We also include a trend and quadratic term in the year of establishment, as well as population-decile fixed effects and fixed effects at the level of sub-districts. In addition, we control for: the predicted share of capitalists within the settlement (Panel A); the predicted share of individuals with previous manufacturing occupations within the settlement (Panel B); the predicted immigration flows (Panel C); the full, actual occupational structure, the share of males who do not speak Hebrew, and the share of illiterate males (Panel D); and the (log) distance between the first-highest and second-highest shares across origins and the share of people born in Palestine (Panel E). Standard errors are clustered at the sub-district level and reported between parentheses. F-statistics are reported between brackets.

B.3 Identification and interpretation of the effect

By construction, our instrument hinges on variation in immigrant flows over time. This variation generates different levels of uniformity across settlements, but also possibly: (i) different, observable settlers' characteristics; and (ii) different ex-ante ideologies, preferences, and beliefs for settlers. In our dichotomy of Section 4, these confounders might alter the *interpretation* of the diversity effect.

Settlers' characteristics Settlements with an exogenously diverse population could differ from other communities of Mandatory Palestine along other characteristics than

just diversity, e.g., skills, occupations, language, and overall migration incidence. To reduce concerns that these mitigating factors drive our main estimates, we consider two alternative strategies. First, we rely on our shift-share prediction strategy to construct other mappings from *immigrants* over time onto *settlers* across space. Letting x_τ denote a characteristic of a migrant wave, we construct,

$$\widetilde{\chi_i} = \sum_\tau \widetilde{s}_{i\tau} \times x_\tau,$$

and we condition our baseline specification on these shift-share predictions, $\widetilde{\chi_i}$. Second, we do observe actual settlers' characteristics around 1945, and we can directly control for these (endogenous, ex-post) factors in our baseline specification.

Table B5 reports the outcome of both strategies, conditioning the baseline specification on: the predicted share of capitalists within the settlement; the predicted share of individuals with previous occupations in manufacturing; the predicted immigration flows; the full, actual occupational structure, the share of males who do not speak Hebrew, and the share of illiterate males; and indicators of group dynamics (the distance between the first-highest and second-highest shares across origins and the share of people born in Palestine). These results indicate that the diversity effect is not due to diverse settlements being more prosperous or producing more human capital or giving less power to a majority group compared to the second-most-prominent group; they also provide little evidence in favor of competition for leadership in heterogeneous societies as being the main channel behind the rise of leaders in diverse communities. We find that neither the inclusion of the former characteristics (industrial structure, literacy or use of Hebrew as a main or secondary language) nor of the latter characteristics (e.g., prevalence of a majority group from a certain origin) affect the correlation between diversity and the rise of leaders.

Settlement types and ideology In Table B6, we test whether the diversity effect could be mitigated by the rise of different ideologies or different settlement structures. While we do find that diversity has different effects across different settlement structures (treatment heterogeneity—see Section 5.1), we do not find that it mitigates our effect. For transparency, we control for the settlement type in Table B6 and report the estimates for the kibbutz and the moshav (against the omitted category including cities and settlement with other structures, e.g., the moshava). The main estimates are left unchanged by these inclusions; we note however that the kibbutz is more likely to see the rise of political leaders (as discussed, e.g., in Rayman, 1981).

Table B6. Diversity, the rise of societal leaders and inter-ethnic relationships—robustness to controlling for settlement type.

	All leaders (1)	Mil. leaders (2)	Pol. leaders (3)	Cul. leaders (4)	Casualties (5)	Pacts (6)
Herfindahl index	-4.234	-1.541	-1.321	-1.370	1.473	-1.169
	(0.851)	(0.397)	(0.497)	(0.442)	(0.471)	(0.475)
Kibbutz	0.264	0.090	0.152	0.021	0.046	0.058
	(0.191)	(0.093)	(0.085)	(0.043)	(0.064)	(0.108)
Moshav	-0.010	-0.001	0.007	-0.016	-0.024	0.108
	(0.273)	(0.109)	(0.122)	(0.051)	(0.049)	(0.123)
Observations	316	316	316	316	316	316
F-stat	71.33	71.33	71.33	71.33	71.33	71.33

Notes: A unit of observation is a settlement. The dependent variable is: the number of leaders having resided in the settlement in column (1); military leaders in column (2); political leaders in column (3); cultural leaders in column (4); the probability of civilian casualties in nearby Arab villages within 10 kms in column (5); and the total number of pacts between each Jewish settlement of our sample and neighboring Arab villages in column (6). All specifications include the control variables used in Table 3; we also control for settlement types (Kibbutz, Moshav, and Other—the omitted category including cities). Standard errors are clustered at the sub-district level and reported between parentheses.

B.4 The selection, allocation, and heterogeneity of leaders

This section discusses specific issues pertaining to the analysis of leaders and their rise across settlements.

Additional evidence using Knesset minutes In panel B of Table 3, we analyze minutes of Knesset sessions (1949–1977) and approximately 20,000 individual interventions to classify leaders by their rhetoric and thematic preferences at the Knesset. Table B7 extends this analysis by examining more themes concerning all interventions (Panel A) and interventions specifically addressing Arab-Jewish relations (Panel B).

In Panel A, columns (1)-(2) classify leaders by their emphasis on security: "Secu. (+)" leaders fall in the top quartile for security-related interventions, while "Secu. (-)" includes the remaining 75%. Columns (3)-(4) apply the same 25%/75% split for economic issues, and columns (5)-(6) use leaders' average importance scores. We find that leaders from diverse settlements are less likely to prioritize security and more likely to focus on economic or cohesion-related issues (see Table 3 for the latter). Second, panel B replicates the dichotomy used in panel B of Table 3. Columns (1)-(2) distinguish between leaders with constructive or optimistic tone; columns (3)-(4) focus on references to community cohesion; and columns (5)-(6) split by positivity scores. While positive rhetoric and cohesion references are more common among leaders from diverse settlements, these effects

Table B7. Diversity and the rise of societal leaders—evidence using Knesset minutes.

Leaders	Secu. (+) (1)	Secu. (-) (2)	Econ. (+) (3)	Econ. (-) (4)	Imp. (+) (5)	Imp. (-) (6)
Panel A: All topics						
Herfindahl index	-1.076	-1.619	-1.399	-1.296	-1.195	-1.500
	(0.459)	(0.410)	(0.387)	(0.442)	(0.442)	(0.387)
Observations	316	316	316	316	316	316
F-stat	55.31	55.31	55.31	55.31	55.31	55.31
Mean dep. var.	0.076	0.182	0.076	0.176	0.085	0.174

Leaders	Tone (+) (1)	Tone (-) (2)	Cohesion (+) (3)	Cohesion (-) (4)	Positivity (+) (5)	Positivity (-) (6)
Panel B: Arab-Jewish relations						
Herfindahl index	-1.351	-1.526	-1.452	-1.486	-1.262	-1.615
	(0.534)	(0.429)	(0.471)	(0.320)	(0.480)	(0.464)
Observations	316	316	316	316	316	316
F-stat	55.31	55.31	55.31	55.31	55.31	55.31
Mean dep. var.	0.071	0.156	0.074	0.168	0.076	0.159

Notes: A unit of observation is a settlement. In panel A, the dependent variable is the number of leaders having resided in the settlement: in the top quartile for mentioning security in column (1), and in the bottom 75% (column 2); in the top quartile for mentioning economic issues in column (3), and in the bottom 75% (column 4); and in the top quartile for the importance score—based on a subjective evaluation of the importance of the discussion—in column (5), and in the bottom 75% (column 6). In panel B, the dependent variable is the number of leaders having resided in the settlement: belonging in the top quartile of leaders in terms of the positive tonality of their interventions at the Knesset (1949–1977, column 1); belonging in the bottom 75% in terms of the positive tonality of their interventions at the Knesset (1949–1977, column 2); belonging in the top quartile of leaders in terms of mentions of community cohesion at the Knesset (1949–1977, column 1); belonging in the bottom 75% in terms of mentions of community cohesion at the Knesset (1949–1977, column 4); belonging in the top quartile of leaders in terms of the positivity score (column 5); and belonging in the bottom 75% in terms of positivity score (column 6). Details on data collection, processing, and description of these interventions are provided in Appendix A.4. The explaining variable is a Herfindahl index of origins across all settlers; the instrument is a measure of uniformity based on (i) the timing of settlement creation as predicted by the "frontier expansion" algorithm and (ii) a Herfindahl index of migrant flows. All specifications include the variables used in the "frontier expansion" algorithm (see Section 3.1): all separate geographic characteristics used to capture suitability; and the initial allocation of settlements and travel time to settlements around 1920. We also include a trend and quadratic term in the year of establishment, as well as population-decile fixed effects and fixed effects at the level of sub-districts. Standard errors are clustered at the sub-district level and reported between parentheses.

are somewhat weaker than those found for intra-ethnic interactions.

The rise of heterogeneous leaders We mention in Section 4.1 results pertaining to the governmental roles, political affiliation(s), and birthplace of leaders. In Table B8, we report the diversity effect on: the rise of leaders with prominent political roles in panel A (Knesset members, columns 1 and 2, governmental roles, columns 3 and 4, books, columns 5 and 6); the rise of leaders from different parties in panel B; and the rise of leaders from different origins in panel C. In general, we find that diverse settlements produce leaders across the board, with the exception of leaders from conservative parties

Table B8. Diversity and the rise of societal leaders—governmental roles, political affiliations, and provenance.

Leaders	Knes. (1)	Knes. (>1) (2)	Govt (3)	Govt (>1) (4)	Pub. (5)	Pub. (>1) (6)
Panel A: Governmental roles and publications						
Herfindahl	-1.541	-1.532	-1.230	-1.265	-1.619	-1.683
	(0.391)	(0.304)	(0.406)	(0.405)	(0.216)	(0.391)
Observations	316	316	316	316	316	316
F-stat	55.31	55.31	55.31	55.31	55.31	55.31
Leaders	HaAvoda (1)	Labor (2)	Herut (3)	Relig. (4)	Zion. (5)	Other (6)
Panel B: Political parties						
Herfindahl	-1.505	-0.605	-1.324	-0.842	-1.181	-0.354
	(0.318)	(0.211)	(0.454)	(0.327)	(0.332)	(0.080)
Observations	316	316	316	316	316	316
F-stat	55.31	55.31	55.31	55.31	55.31	55.31
Leaders	Pales. (1)	Poland (2)	Russia (3)	Ukr/Bel. (4)	Europe (5)	Mid-East (6)
Panel C: Origins						
Herfindahl	-0.767	-1.460	-1.647	-1.888	-1.255	0.055
	(0.275)	(0.480)	(0.426)	(0.376)	(0.363)	(0.090)
Observations	316	316	316	316	316	316
F-stat	55.31	55.31	55.31	55.31	55.31	55.31

Notes: A unit of observation is a settlement. The dependent variables are the number of leaders having resided in the settlement: depending on their governmental roles in panel A; depending on their political affiliation in panel B; and depending on their origins in panel C. The explaining variable is a Herfindahl index of origins across all settlers; the instrument is a measure of uniformity based on (i) the timing of settlement creation as predicted by the "frontier expansion" algorithm and (ii) a Herfindahl index of migrant flows. All specifications include the variables used in the "frontier expansion" algorithm (see Section 3.1): all separate geographic characteristics used to capture suitability; and the initial allocation of settlements and travel time to settlements around 1920. We also include a trend and quadratic term in the year of establishment, as well as population-decile fixed effects and fixed effects at the level of sub-districts. Standard errors are clustered at the sub-district level and reported between parentheses.

(panel B of Table B8), from religious parties (to a lesser extent, panel B), or originating from the Middle-East (to a lesser extent, panel C).

The rise of influential leaders We finally construct standardized leader "scores" based on the number of governmental roles and spells at the Knesset (column 1 of Table B9), the length of their biographies (column 2), and the number of publications (column 3). We find that a 0.10 higher Herfindahl index: reduces the governmental score by 0.28 standard deviations; reduces the biography score by 0.11 standard deviations; and reduces the publication score by 0.30 standard deviations. Overall, this evidence does

Table B9. Diversity and the rise of societal leaders—scores.

Leaders	Political (1)	Description (2)	Publications (3)
Herfindahl index	-2.850	-1.078	-2.993
	(0.779)	(0.390)	(1.004)
Observations	316	316	316
F-stat	55.31	55.31	55.31

Notes: A unit of observation is a settlement. The dependent variable is the number of leaders having resided in the settlement: a standardized leader "score" based on the number of governmental roles and spells at the Knesset (column 1); a standardized leader "score" based on the length of their biographies (column 2); and a standardized leader "score" based on the number of publications (column 3). The explaining variable is a Herfindahl index of origins across all settlers; the instrument is a measure of uniformity based on (i) the timing of settlement creation as predicted by the "frontier expansion" algorithm and (ii) a Herfindahl index of migrant flows. All specifications include the variables used in the "frontier expansion" algorithm (see Section 3.1): all separate geographic characteristics used to capture suitability; and the initial allocation of settlements and travel time to settlements around 1920. We also include a trend and quadratic term in the year of establishment, as well as population-decile fixed effects and fixed effects at the level of sub-districts. Standard errors are clustered at the sub-district level and reported between parentheses.

not provide more insight than our baseline results and rather constitute a robustness check—showing that our findings are not driven by "unimportant" leaders.

Table B10. Diversity and the rise of societal leaders—alternative data source (Encyclopedia of the Founders and Builders of Israel).

Founders	All (1)	Influential (2)
Herfindahl index	-45.195	-8.384
	(17.974)	(3.489)
Observations	316	316
F-stat	55.31	55.31

Notes: A unit of observation is a settlement. The dependent variable is the number of founders (column 1) and influential founders (column 2) having resided in the settlement, both constructed from the Encyclopedia of the Founders and Builders of Israel. The explaining variable is a Herfindahl index of origins across all settlers; the instrument is a shift-share prediction based on the timing of settlement creation as predicted by the "frontier expansion" algorithm and a Herfindahl index of yearly migrant flows. All specifications include the variables used in the "frontier expansion" algorithm (see Section 3.1): all separate geographic characteristics used to capture suitability; and the initial allocation of settlements and travel time to settlements around 1920. We also include a trend and quadratic term in the year of establishment, as well as population-decile fixed effects and fixed effects at the level of sub-districts. Standard errors are clustered at the sub-district level and reported between parentheses.

An external encyclopedia of influential settlers We now draw on another data source, which markedly differs from our baseline dataset. The construction of our baseline dataset induces some discretion in who to include or not, but such a selected biographical approach has the advantage of providing extensive information about these selected leaders. In this alternative exercise, we rather rely on externally-curated data from the Encyclopedia of the Founders and Builders of Israel. The selection is then ex-

ternal, and with a lower threshold: we are able to properly locate about 5,150 "influential settlers" from this source. The disadvantage is that the information is more limited. We however collect places of origin, residence, birth information and aliyah date, as well as some information about their occupations and an indicator of importance.[43] Given the less extensive information available in this encyclopedia, we use a probabilistic allocation of these influential settlers across their documented locations (e.g., we would allocate 0.5 individuals to locations A and B if one individual is reported to have resided in both places)—in contrast with our baseline procedure described in Section 2.

In Table B10, we replicate Table 3 with this alternative source, using 5,150 individuals (column 1) or a selection of 1,300 "leaders" identified as being more important by the large language model (column 2). The estimate in column (2) is about a fifth of the estimate in column (1), reflecting that the selection of important versus less important leaders is probably quite similar across settlements.

Alternative allocation of leaders Finally, we probe the robustness of our findings to alternative allocations of leaders across settlements. More specifically, we split leaders across settlements in a probabilistic manner, depending on their different residences in Panel A of Table B11; we remove leaders with more than one residence in Panel B; and we divide leaders by their length of stay in Panel C (more than five years versus less than five years). If anything, splitting leaders across settlements tends to increase our estimate of interest, when removing mobile leaders does not modify our qualitative insight. We do find that the diversity effect is strong for leaders having stayed more than five years within their assigned settlement. Overall, this exercise reduces concerns about the discretionary nature of our assignment procedure to settlements.

In Panels C and D, we also look at other measures of leadership: (i) pioneers with some responsibilities within the Yishuv against non-pioneers (Panel C), (ii) leaders from the local majority group (e.g., leaders from Eastern Europe when the larger group of settlers is from Eastern Europe, see Panel D), (iii) religious leaders (Panel D), (iv) civil leaders (e.g., heading the Central Bank of Israel, Panel D), or (v) female leaders (Panel D).

B.5 War and inter-ethnic relationships

Alternative outcomes In this section, we consider alternative outcomes for our study of war and inter-ethnic relationships (as a complement to Section 4.2). In panel A of Table 4, we choose a certain radius to define exposure to the 1948 Arab-Israeli War. In

[43] We perform this task using a large language model, and we check the output using human research assistance. The model performs very well; however, the indicator of importance might be subject to discussions. We have also used an indicator of importance based on the length of the biography with similar results.

Table B11. Diversity and the rise of societal leaders—alternative allocation of leaders.

	All	Military	Political	Cultural
Leaders	(1)	(2)	(3)	(4)
Panel A: Splitting leaders across settlements				
Herfindahl	-6.476	-2.133	-2.020	-2.322
	(1.042)	(0.404)	(0.659)	(0.635)
Observations	316	316	316	316
F-stat	55.31	55.31	55.31	55.31
	All	Military	Political	Cultural
Leaders	(1)	(2)	(3)	(4)
Panel B: Removing mobile leaders				
Herfindahl	-1.817	-0.386	-0.581	-1.044
	(0.560)	(0.239)	(0.355)	(0.179)
Observations	316	316	316	316
F-stat	55.31	55.31	55.31	55.31
	Short stay	Long stay	Non-pioneer	Pioneer
Leaders	(1)	(2)	(3)	(4)
Panel C: Pioneers, length of stay				
Herfindahl	-1.199	-4.826	-1.227	-1.492
	(0.437)	(0.937)	(0.559)	(0.500)
Observations	316	316	316	316
F-stat	55.31	55.31	55.31	55.31
	Dominant group	Religious	Civil	Female
Leaders	(1)	(2)	(3)	(4)
Panel D: Dominant group, other leader types, and gender				
Herfindahl	-0.313	-0.484	-1.316	-0.348
	(0.259)	(0.221)	(0.517)	(0.155)
Observations	316	316	316	316
F-stat	55.31	55.31	55.31	55.31

Notes: A unit of observation is a settlement. In Panels A and B, the dependent variable is the number of leaders having resided in the settlement: military, political, and cultural leaders in column (1); military leaders in column (2); political leaders in column (3); and cultural leaders in column (4). In Panel C (columns 1, 2), we divide leaders by their length of stay (more than five years versus less than five years). In Panel C (columns 3, 4), the dependent variables are the number of non-pioneer/pioneer leaders having resided in the settlement. In Panel D, we look at leaders from the local majority group (e.g., leaders from Eastern Europe when the larger group of settlers is from Eastern Europe), religious leaders, civil leaders (e.g., Governor of the Bank of Israel), and female leaders. The explaining variable is a Herfindahl index of origins across all settlers; the instrument is a measure of uniformity based on (i) the timing of settlement creation as predicted by the "frontier expansion" algorithm and (ii) a Herfindahl index of migrant flows. All specifications include the variables used in Table 3. Standard errors are clustered at the sub-district level and reported between parentheses.

panel A of Table B12, we construct (inverse) distance-weighted outcomes for the same set of events and show that diversity markedly reduces the likelihood to be in the proximity of Arab civilian casualties (as in our main Table 4). In panel B, we focus on this civilian casualties-diversity relationship and vary the exposure radius. We find some effect when

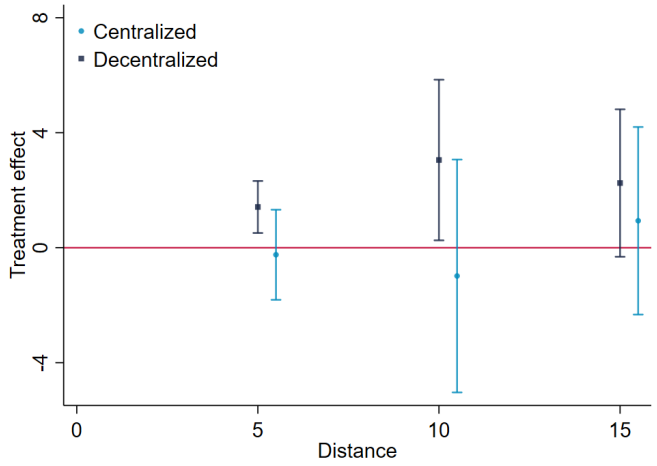
Table B12. Diversity and inter-ethnic relationships—alternative measurement.

War	Depopulated (1)	Attack (2)	Civilian casualties (3)
Panel A: Distance-weighted probabilities			
Herfindahl index	0.025	0.026	2.038
	(0.268)	(0.283)	(0.553)
Observations	316	316	316
F-stat	70.48	70.48	70.48
	Within 5 kms	Within 15 kms	Within 20 kms
Civilian casualties	(1)	(2)	(3)
Panel B: Civilian casualties in nearby Arab villages			
Herfindahl index	0.450	0.654	0.331
	(0.471)	(0.306)	(0.194)
Observations	316	316	316
F-stat	70.48	70.48	70.48
	All types	Prevention	Cooperation
Pacts	(1)	(2)	(3)
Panel C: Pacts (all types, and alternative types)			
Herfindahl index	-1.127	-0.217	0.001
	(0.446)	(0.196)	(0.237)
Observations	316	316	316
F-stat	55.31	55.31	55.31

Notes: A unit of observation is a settlement. In panel A, the dependent variables are distance-weighted probability to be exposed to the following events involving Arab villages: depopulated in column (1); attacked in column (2); and having civilian casualties in column (3). In panel B, the dependent variable is the probability of civilian casualties in nearby Arab villages within 5 (column 1), 15 (column 2) and 20 kms (column 3). In panel C, the dependent variable is the number of pacts, the number of cooperation pacts, and the measure of prevention pacts between each Jewish settlement of our sample and neighboring Arab villages. The explaining variable is a Herfindahl index of origins across all settlers; the instrument is a measure of uniformity based on (i) the timing of settlement creation as predicted by the "frontier expansion" algorithm and (ii) a Herfindahl index of migrant flows. All specifications include the variables used in the "frontier expansion" algorithm (see Section 3.1): all separate geographic characteristics used to capture suitability; and the initial allocation of settlements and travel time to settlements around 1920. We also include a trend and quadratic term in the year of establishment, as well as population-decile fixed effects and fixed effects at the level of sub-districts. Finally, we control for the Arab population and the presence of military Arab troops in these nearby villages (in panels A and B). Standard errors are clustered at the sub-district level and reported between parentheses.

we include nearby villages within 5, 15 or 20 kilometers, but the effect is non-linear: it is lower for a 5-km radius, reaches a maximum for a 10-km radius and then gradually fades away. In panel C, we rather focus on our other indicators of inter-ethnic relationships: the informal agreements between Jewish and Arab settlements (Azoulay, 2014). We then sum all pacts, contrary to panel B of Table 4; we also consider separately the pacts classified as cooperation (e.g., declaration of solidarity) or prevention of a conflict (e.g., a warning about the arrival of troops).

Figure B5. War and inter-ethnic relationships—the role of centralization.



Notes: This Figure reports a sensitivity analysis for the "casualties" estimate (column 3, panel A of Table 4). The reported estimates correspond to separate regressions with: (i) dummies equal to 1 if there are civilian casualties within 5, 10, 15 kilometers for events that were initiated and carried out in a centralized manner (light blue); (ii) dummies equal to 1 if there are civilian casualties within 5, 10, 15 kilometers for events that were initiated and carried out in a decentralized manner (dark blue).

The role of centralization and local communities Most operations behind our analysis of Table 4 (panel A) were initiated and monitored in a centralized manner, e.g., through the use of mobile regional infantry brigades with limited local ties—a feature that is better discussed in Section 1.4. To provide support behind our local interpretation of inter-ethnic relationships during the war, we report a decomposition of the "casualties" estimate (column 3, panel A of Table 4) and contrast events that were initiated and carried out in a centralized manner versus those initiated more locally. Figure B5 shows that our estimate indeed relies on the less frequent, locally-motivated events.

B.6 The heterogeneity across settlements

In this section, we shed some light on survival biases for certain of our data sources, on the (endogenous) nature of settlement organization, and on how kibbutzim might differ from other settlements.

The heterogeneity across settlements Our baseline specification includes 316 settlements. Our analysis however partly relies on: a peculiar settlement structure (the kibbutz); a selection of settlements covered by meeting minutes; and a selection of settlements covered by photographs. We shed light onto the characteristics of these different settlements in Table B13: (A) their actual and predicted uniformity; (B) selected geographic variables; (C) proximity to transportation infrastructure; (D) proximity to initial settlements; (E) settlement creation; (F) general composition (e.g., share of migrants from

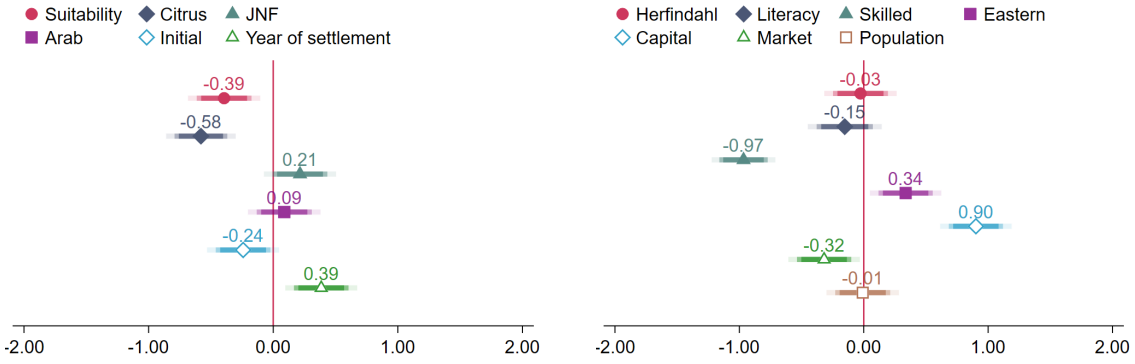
Table B13. The heterogeneity across settlements—all settlements, kibbutzim, settlements with retrieved minutes, and settlements with retrieved photographs.

Sample	All settlements		Kibbutzim	Minutes	Photos
	Mean	Std dev.	Mean	Mean	Mean
Panel A: Uniformity within settlements					
Herfindahl index (normalized)	0.348	0.248	0.342	0.265	0.223
Predicted uniformity (normalized)	0.404	0.237	0.360	0.463	0.468
Panel B: Geography					
Latitude	32.362	0.473	32.363	32.403	32.456
Longitude	35.086	0.306	35.142	35.144	35.184
Elevation	84.304	168.102	79.801	64.591	75.579
Slope	3.424	4.137	3.555	3.686	3.950
Average precipitation	42.133	8.195	39.729	41.168	42.295
Panel C: Transportation infrastructure					
Inverse distance to railways	0.307	0.393	0.312	0.404	0.432
Inverse distance to roads	0.778	0.399	0.687	0.842	0.937
Panel D: Proximity to initial settlements					
Proximity to Jewish settlements	0.099	0.024	0.095	0.105	0.107
Proximity to Arab settlements	0.242	0.121	0.249	0.263	0.294
Panel E: Year of establishment					
Median year of settlement creation	1935	22	1938	1925	1921
Panel F: Composition					
Settler share from Eastern Europe	0.418	0.229	0.412	0.477	0.509
Settler share from Western Europe	0.264	0.209	0.299	0.228	0.180
Settler share from the Balkans	0.199	0.181	0.205	0.196	0.196
Settler share from the Maghreb	0.020	0.082	0.014	0.012	0.017
Settler share from the Middle-East	0.073	0.156	0.036	0.061	0.069
Panel G: Settlement organization					
Kibbutz	0.424	0.495	1.000	0.542	0.395
Observations	316		134	107	43

Notes: This table displays descriptive statistics for all settlements in columns (1) and (2), kibbutzim in column (3), the selection of settlements covered by meeting minutes in column (4), and the selection of settlements covered by photographs in column (5).

Eastern Europe); and (G) settlement organization. In general, kibbutzim are quite similar to the average settlement, except for their slightly higher share of migrants from Western Europe and lower share of migrants from the Middle-East. By contrast, settlements covered by minutes or photographs are quite different from the average settlement: they are older, better connected to the transportation infrastructure, and in denser areas (of Jewish and Arab settlements).

Figure B6. Settlement type and settlement characteristics.



(a) Kibbutzim (1/2)

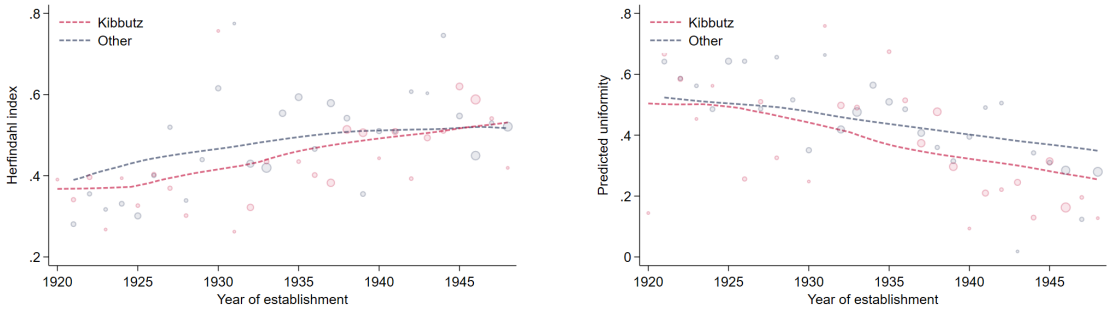
(b) Kibbutzim (2/2)

Notes: This Figure reports the standardized difference between kibbutzim and other settlement types in: the suitability index, the predicted yield for citrus, the land share owned by the Jewish National Funds, the (local) density of Arab settlements, the proximity to initial Jewish settlements, the year of settlement (panel a); the normalized Herfindahl index across origins, the share of settlers who are literate, the share of white collar occupations, the standardized share of migrants from Eastern Europe, a standardized measure of capital in 1945, a standardized measure of market access to other Jewish settlements in 1947, and a standardized measure of initial settlement population.

The (endogenous) nature of settlement organization We provide a (conditional) balance test in Figure B6, in which we derive the difference in standardized outcomes between kibbutzim and other structures. These (standardized) outcomes are: the suitability index, the predicted yield for citrus, the land share owned by the Jewish National Funds, the (local) density of Arab settlements, the proximity to initial Jewish settlements, the year of settlement (panel a); the normalized Herfindahl index across origins, the share of settlers who are literate, the share of white collar occupations, the standardized share of migrants from Eastern Europe, a standardized measure of capital in 1945, a standardized measure of market access to other Jewish settlements in 1947, and a standardized measure of initial settlement population. We find that kibbutzim are in less suitable locations, tend to be created later than other settlements, are (far) less likely to host white-collar workers, are more likely to host migrants from Eastern Europe, are better endowed in capital, and are less well connected to other settlements. They are however equally diverse. If anything, kibbutzim do not seem more prone to the rise of leaders from the perspective of *these* observables.

We plot actual diversity across different years of settlement and settlement type in panel (a) of Figure B7. Settlements become less diverse as their date of creation increases, and these effects are linear and not different across settlement types. This finding rationalizes controlling for trends in the date of settlement creation, but does not show any strong differential dynamics between kibbutzim and other settlements. We take this evidence as suggestive that the "kibbutz" premium to the diversity effect has to do with the

Figure B7. Settlement type, year of establishment and diversity.



(a) Herfindahl index (b) Predicted uniformity

Notes: Panel (a) plots the average Herfindahl index across years of establishment (x-axis) for kibbutzim (red) and other types of settlements (blue). Panel (b) plots the average shift-share prediction of uniformity across years of establishment (x-axis) for kibbutzim (red) and other types of settlements (blue)

organization of life within the kibbutz more than their composition—an insight that we provide with a more qualitative approach in Section 5.2.

Table B14. Treatment heterogeneity—robustness to confounders.

	All leaders	Mil. lead.	Pol. lead.	Cul. lead.	Casualties	Pacts
	(1)	(2)	(3)	(4)	(5)	(6)
Herfindahl	-2.612	-1.039	-0.445	-1.129	1.358	-0.393
	(1.253)	(0.479)	(0.566)	(0.410)	(0.608)	(0.575)
Herfindahl × kibbutz	-7.025	-1.026	-5.185	-0.815	1.335	-2.890
	(5.112)	(2.195)	(2.369)	(1.149)	(2.422)	(1.389)
Observations	316	316	316	316	316	316
F-stat (Herfindahl)	42.93	42.93	42.93	42.93	42.93	42.93
F-stat (Interaction)	8.16	8.16	8.16	8.16	8.16	8.16

Notes: A unit of observation is a settlement. The dependent variable is: the number of leaders having resided in the settlement in column (1); military leaders in column (2); political leaders in column (3); cultural leaders in column (4); the probability of civilian casualties in nearby Arab villages within 10 kms in column (5); and the total number of pacts between each Jewish settlement of our sample and neighboring Arab villages in column (6). All specifications include the same controls as in Table 5, with the addition of the standardized share of migrants from Eastern Europe, the standardized share of white collar occupations, and the standardized share of farmers. F-statistics are derived using the weak instrument F-test proposed in Sanderson and Windmeijer (2016). Standard errors, in parentheses, are clustered at the sub-district level.

Finally, the previously-documented differences in settlement characteristics might bias the treatment heterogeneity estimated in Section 5.1. In Table B14, we reproduce Table 5, but conditioning on the incidence of migrants from Eastern Europe and on local skills (i.e., the standardized share of white collar occupations and the standardized share of farmers). While the resulting estimates are slightly noisier, the qualitative insights of Table B14 remain unchanged. In unreported checks, we also verified that these estimates are robust to interacting diversity with each of these characteristics separately—in a

specification with three endogenous variables and three instruments.

Table B15. Diversity and nation-building within the kibbutz—other topics.

Topics	Immigration (1)	Economy (2)	Land (3)	Security (4)	Youth (5)
Herfindahl index	0.270	0.047	-0.072	-0.092	0.086
	(0.150)	(0.230)	(0.151)	(0.059)	(0.069)
Observations	316	316	316	316	316
F-stat	55.31	55.31	55.31	55.31	55.31

Notes: A unit of observation is a settlement. The explaining variable is a Herfindahl index of origins across all settlers; the instrument is a measure of uniformity based on (i) the timing of settlement creation as predicted by the "frontier expansion" algorithm and (ii) a Herfindahl index of migrant flows. All specifications include the variables used in the "frontier expansion" algorithm (see Section 3.1): all separate geographic characteristics used to capture suitability; and the initial allocation of settlements and travel time to settlements around 1920. We also include a trend and quadratic term in the year of establishment, as well as population-decile fixed effects and fixed effects at the level of sub-districts. Standard errors are clustered at the sub-district level and reported between parentheses.

B.7 Community cohesion within the kibbutz

In this final section, we provide complements to Section 5. Section 5 relies on a topic analysis of meeting minutes within the kibbutz, focusing on their discussion of community cohesion (Table 6), the nature of "challenging" discussions (i.e., the entries associated with a negative tone, see Figure 9), and general statistics about the frequency of certain tones and subjects. In Appendix Table B15, we shed light on the average differences in topic coverage between less and more diverse settlements, irrespective of the tone. Diverse kibbutzim discuss less the absorption of immigrants, and slightly more land and security topics.

In Section 5, we mention results pertaining to the rise of different leaders in the kibbutz (their governmental roles, political affiliation(s), and birthplaces). In Appendix Table B16, we report the estimates of Equation (6) for these outcomes: we find that the diversity effect on the rise of prominent leaders is confined to kibbutzim, except for religious leaders, conservative leaders and leaders born in Palestine or originally from the Middle-East.

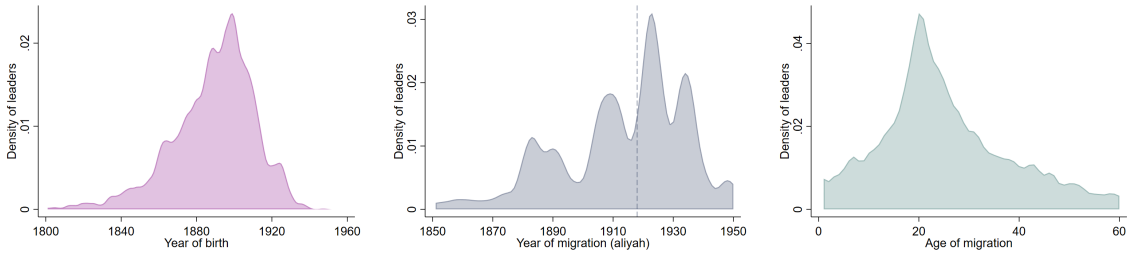
Table B16. Diversity and the rise of different leaders in the kibbutz.

Leaders	Knes. (1)	Knes. (>1) (2)	Govt (3)	Govt (>1) (4)	Pub. (5)	Pub. (>1) (6)
Panel A: Governmental roles and publications						
Herfindahl	-0.526	-0.406	-0.408	-0.709	-0.946	-0.987
	(0.604)	(0.720)	(0.718)	(0.497)	(0.489)	(0.566)
Herfindahl × kibbutz	-5.203	-6.168	-5.561	-3.392	-3.436	-3.922
	(2.171)	(1.881)	(1.946)	(1.720)	(1.850)	(1.836)
Observations	316	316	316	316	316	316
Leaders	HaAvoda (1)	Labor (2)	Herut (3)	Relig. (4)	Zion. (5)	Other (6)
Panel B: Political parties						
Herfindahl	-0.593	0.006	-0.881	-0.627	-0.743	-0.271
	(0.633)	(0.454)	(0.533)	(0.301)	(0.439)	(0.123)
Herfindahl × kibbutz	-4.869	-3.886	-2.312	-1.147	-2.398	-0.053
	(2.106)	(1.599)	(1.068)	(0.908)	(1.738)	(0.330)
Observations	316	316	316	316	316	316
Leaders	Pales. (1)	Poland (2)	Russia (3)	Ukr/Bel. (4)	Europe (5)	Mid-east (6)
Panel C: Origins						
Herfindahl	-0.483	-0.882	-1.037	-1.024	-0.626	0.037
	(0.330)	(0.522)	(0.587)	(0.765)	(0.499)	(0.061)
Herfindahl × kibbutz	-0.329	-2.727	-3.380	-4.477	-3.216	0.063
	(0.796)	(1.653)	(1.861)	(2.013)	(2.381)	(0.482)
Observations	316	316	316	316	316	316

Notes: A unit of observation is a settlement. The dependent variables are the number of leaders having resided in the settlement: depending on their governmental roles in panel A; depending on their political affiliation in panel B; and depending on their origins in panel C (in the manner of Table B8). The empirical specification is similar to the one displayed in Table 5. Standard errors are clustered at the sub-district level and reported between parentheses.

C Selection and allocation of leaders

Figure C1. Additional descriptives about the "Founders and Builders of Israel".



(a) Birth date (b) Migration year (c) Migration "age"

Notes: This Figure displays the distributions of year at birth (panel a), year of migration (panel b), and age during the aliyah (panel c) for the larger, external sample of about 6,000 leaders extracted from the *Encyclopedia of the Founders and Builders of Israel* (Tidhar, 1947).

C.1 Additional descriptives about the "Founders and Builders of Israel"

As mentioned in Section 2, we exploit the 19 volumes of the *Encyclopedia of the Founders and Builders of Israel* curated by Tidhar (1947) to generate a less arbitrary selection of leaders (which we use in a robustness check, see Appendix B). Another advantage of using this data source is that we retrieve basic information on the birth location and birth date, as well as the migration year and the age at the aliyah (within the sample of leaders born outside of Palestine). We report these descriptive statistics in Figure C1: (a) most leaders are born between 1880 and 1920, inducing an age at the time of the creation of the state between 28 and 68; (b) there are non-negligible migration waves prior to our period of interest for those founders (these earlier migration waves are far less present in our baseline sample, the reason for this discrepancy is that Tidhar (1947) gives more prominence to the early precursors); and (c) the age during the aliyah peaks around 20-25 years old, meaning that the first location of residence coincides with the impressionable years.

C.2 Our baseline selection of leaders

We collect individual-level data on about 240 societal leaders engaged in the process of Israel's nation-building and having lived in a Jewish settlement of Mandatory Palestine. Specifically, we systematically retrieve data from biographical information on: their place and date of birth, their education, the year of migration if relevant, the places of residence in Palestine, the date and place of military service (if relevant), and a set

of later-life outcomes. As regards the latter, we collect the type and length of government or military role covered (Knesset member, government role, Haganah commander, etc.), but also their successive political affiliations and their bibliographies (if relevant). We rely on the following websites to compile our dataset: the Knesset (the unicameral legislature of Israel), Israel Defense Forces, the Jewish Virtual Library, and the Jewish Agency. We complement these sources with the previously-described 19-volume *Encyclopedia of the Founders and Builders of Israel* curated by Tidhar (1947), accessed at `https://www.tidhar.tourolib.org/` and the following (auto) biographies: Ben-Zvi, Rahel Yanait, "Coming Home" (1963); Dayan, Moshe, "Moshe Dayan: Story of my life" (1976); Bondy, Ruth, "The Emissary: A Life of Enzo Sereni" (1976, about Enzo Sereni); Podda, Stefania, "Golda Meir, Grandi donne della storia" (2021, about Golda Meir); Fraser, Thomas G, "Chaim Weizmann: The Zionist Dream" (2010, about Chaim Weizmann); Shilon, Avi, "Menachem Begin" (2012, about Menachem Begin); Avner, Yehuda, "The Prime Ministers: An Intimate Narrative of Israeli Leadership" (2010, about Yitzhak Rabin and Levi Eshkol).

For simplicity, we consider that a leader is a political leader if they assume any strategic political roles at any point of their career (e.g., being a Knesset member or a government member). We label writers (novels, essays, poems, theater etc.), composers or songwriters, film directors, and journalists/editors as cultural leaders. One example is Rachel Bluwstein, an early settler and "pioneer" poetess who resided in kibbutz Degania Alef. We also collect information on other civil and religious figures, e.g., David Horowitz (founder and first Governor of the Bank of Israel who resided in kibbutz Beit Alfa) or Avrohom Yeshaya Karelitz (an influential rabbi who resided in Bnei Brak after migrating in 1933). We however do not use those other civil and religious leaders in our baseline specification. Note that, in principle, our leadership categories (military, political, cultural) are not exclusive, and a few key leaders might tick more than one box. The full list of leaders used in our baseline sample is reported in Table C1 at the end of this Appendix. Finally, and as mentioned in the text already, this work benefited from *outstanding* research assistance by Susan Staszewski.

Table C1: Leaders characteristics

Name	Role	Country of birth	Year of birth	Year of *aliyah*	Allocated settlement	Other places of residence
Aaron Aaronsohn	m	Romania	1876	1882	Zikhron Yaakov	
Abba Eban	p	South Africa	1915	1959	Tel Aviv	
Abba Hushi	m	Poland	1898	1920	Beit Alfa	Nahalal, Haifa
Abba Kovner	m	Lithuania	1918	1945	Ein Hahoresh	
Abraham Meshorer	m	Belarus	1907	1923	Jerusalem	
Abraham Shapira	m	Palestine	1870	-	Jerusalem	
Agnon Shmuel Yosef	c	Austria-Hungary	1888	1908	Tel Aviv	Jerusalem

Continued on next page

Table C1: Leaders characteristics (Continued)

Agron Gershon	c, p	Russian Empire	1893	1924	Jerusalem	
Ahad Ha'am	c	Russian Empire	1856	1922	Tel Aviv	Jerusalem
Aharon Becker	p	Belarus	1905	1925	Petah Tikva	Ramat Gan, Tel Aviv
Aharon Doron	m	Germany	1922	1939	Yagur	
Aharon Goldstein	m	Russia	1902	1921	Givath Hayim	
Aharon Hoter Yishai	m	Russia	1905	1913	Tel Aviv	Hadera, Haifa
Aharon Remez	p	Palestine	1919	-	Tel Aviv	
Aharon Yadlin	p	Palestine	1926	-	Kfar Marmorek	Ben Shemen, Rehovot, Haifa Geniger, Benyamina , Gedera
Akiva Govrin	p	Ukraine	1902	1922	Jerusalem	
Aloni Shulamit	c, p	Poland	1927	1930	Tel Aviv	
Ami Assaf	p	Palestine	1903	-	Rosh Pina	Kfar Yehoshua
Amichai Yehuda	c	Germany	1924	1935	Petah Tikva	Jerusalem, Haifa
Amnon Inbar	m	Poland	1926	1934	Tel Aviv	Ramat Gan
Amos Degani	p	Palestine	1926	-	Kefar Vitkin	Tel Aviv
Argov Sacha	c	Russian Empire	1914	1934	Tel Aviv	
Ari Ankorion	p	Lithuania	1908	1933	Jerusalem	Tel Aviv
Ariel Sharon	p	Palestine	1928	-	Kfar Malal	
Aryeh Ben-Eliezer	p	Poland	1913	1921	Tel Aviv	
Aryeh Disenchik	c	Russian Empire	1907	1934	Tel Aviv	
Aryeh Eliav	c, p	Soviet Russia	1921	1924	Tel Aviv	
Aryeh Sheftel	p	Russia	1905	1947	Rishon Lezion	
Aura Herzog	m	Egypt	1924	1946	Jerusalem	
Avraham Abbas	p	Syria	1912	1929	Kfar Giladi	
Avraham Adan	m	Palestine	1926	-	Kfar Giladi	Rishon LeZion
Avraham Bar-Uryan	m	Romania	1910	1930	Kfar Ganim	Rehovoth Gush Tel Mond
Avraham Drori	p	Poland	1919	1935	Tel Aviv	
Avraham Granot	p	Russia	1890	1907	Tel Aviv	
Avraham Kalfon	p	Palestine	1900	-	Tiberias	Haifa
Avraham Krinitzi	p, m	Belarus	1886	1905	Ramat Gan	Jaffa, Petah Tikva, Tel Aviv
Avraham Shlonsky	c	Russian Empire	1900	1921	Ein Harod	Tel Aviv
Avraham Taviv	p	Yemen	1889	1908	Rishon Lezion	
Bar-Ilan Meir	c	Russian Empire	1880	1923	Jerusalem	
Baruch Kamin	p	Romania	1914	1939	Nirim	
Baruch Azanja	p	Russia	1905	1933	Givath Hayim (Kibbutz)	
Bechor-Shalom Sheetrit	p	Palestine	1895	-	Tiberias	
Belkind Israel	c	Russian Empire	1861	1882	Mikveh Israel	Rishon Lezion, Gdera, Tel Aviv, Jerusalem
Ben Ami Pachter	m	Poland	1919	1925	Nahariya	
Ben-Zion Meir Hai Uziel	c	Palestine	1880	-	Jerusalem	Tel Aviv
Benjamin Mintz	p	Poland	1903	1925	Tel Aviv	
Berl Katznelson	c, p	Russian Empire	1887	1909	Ein Ganim	Hadera, Sejera, Degania Kinneret, Atarot
Bialik Haim Nahman	c	Russian Empire	1837	1924	Tel Aviv	
Bineth Meir Max	c, m	Hungary	1917	1935	Kfar Hassidim	Tel Aviv, Hadera, Rishon LeZion
Binyamin Gibli	m	Palestine	1919	-	Petah Tikva	Ein Ganim
Buber Martin	c	Austria	1878	1938	Jerusalem	
Chaim Ariav	p	Russia	1895	1912	Tel Aviv	Jerusalem
Chaim Arlosoroff	c, p	Russian Empire	1899	1924	Tel Aviv	
Chaim Herzog	m	Ireland	1918	1935	Jerusalem	
Chaim Miller	m	Austria	1921	1939	Raanana	
Chaim Weizmann	p	Belarus	1874	1918	Kfar Marmorek	Rehovoth
Chanan Rapaport	p	Ukraine	1928	1935	Nesher	Jerusalem
Dan Even	p	Germany	1912	1934	Givath Hayim (Kibbutz)	Haifa, Herzlyia

Continued on next page

Table C1: Leaders characteristics (Continued)

Dan Tolkowsky	c, m	Palestine	1921	-	Tel Aviv	
David Bar-Rav-Hai	m	Ukraine	1894	1924	Haifa	
David Ben-Gurion	p	Poland	1886	1906	Petah Tikva	Sejera, Zichron Yaakov, Tel Aviv
David Fromer	m	Poland	1920	1936	Tel Aviv	Nahalath Yitzhak Haifa
David Hacohen	p	Russiam Empire	1898	1907	Tel Aviv	
David Horowitz	c	Austria-Hungary	1899	1920	Bet Alfa	Jerusalem, Tel Aviv
David Livschitz	p	Russian Empire	1897	1938	Tel Aviv	
David Remez	p, m	Russian Empire	1886	1913	Beer Touvya	Zichron Ya'akov, Tel-Aviv
David Shaltiel	m	Germany	1903	1923	Jerusalem	Haifa
David Shimoni	c	Russian Empire	1891	1920	Rehovot	Tel Aviv
David-Zvi Pinkas	p	Hungary	1897	1925	Tel Aviv	
Dinitz Simcha	c	Palestine	1929	-	Tel Aviv	
Dizengoff Meir	c, p	Bessarabia	1861	1905	Tel Aviv	Tel Aviv
Dov Yosef	p, m	Canada	1899	1918	Jerusalem	
Dov Bar-Nir	p, m	Belgium	1911	1932	Ein Hahoresh	
Dov Hoz	m	Belarus	1894	1906	Tel Aviv	Degania
Edmond Wilhelm Brillant	m	Poland	1916	1936	Haifa	
Ehud Avriel	p	Austria	1917	1940	Naot Mordehai	
Ehud Barak	p, m	Palestine	1942	-	Mishmar Hasharon	
Ehud Olmert	p, m	Palestine	1945	-	Jerusalem	
Elazar David	c, m	Yugoslavia	1925	1940	Kibbutz Ein Shemer	
Eliezer Ben-Yehuda	c	Russian Empire	1858	1881	Jerusalem	
Eliezer Kaplan	p, m	Belarus	1891	1920	Tel Aviv	
Eliezer Shostak	p	Poland	1911	1935	Herzliya	
Eliyahu Golomb	m	Belarus	1893	1909	Tel Aviv	Ramat Gan
Eliyahu Hacarmeli	p, m	Palestine	1891	-	Haifa	Tiberias, Jerusalem
Eliyahu Mazur	p, m	Poland	1889	1940	Tel Aviv	
Emil Habibi	p	Palestine	1922	-	Haifa	
Enzo Sereni	m	Italy	1905	1927	Givath Brenner	
Ephraim Katzir	p	Ukraine	1916	1925	Jerusalem	
Esther Cailingold	m	Great Britain	1925	1946	Jerusalem	
Esther Raab	c	Palestine	1894	-	Degania Alef	Petah Tikva, Tel Aviv
Finkel Shimon	c	Russian Empire	1905	1924	Tel Aviv	
Frumkin Rabbi Arieh Leib	c	Lithuania	1845	1871	Petah Tikva	
Gershom Schocken	c	Germany	1912	1933	Tel Aviv	
Gideon Ben-Israel	p, m	Palestine	1923	-	Haifa	
Gideon Shocken	m	Germany	1919	1934	Jerusalem	
Giora Yoseftal	p	Germany	1912	1938	Givath Hayim (Kibbutz)	Gal Ed
Goel Levitzky	m	Palestine	1917	-	Ein Ganim	Alonim
Golda Meir	p	Ukraine	1898	1921	Merhavya (Kibbutz)	Tel Aviv, Jerusalem
Goldberg Lea	c	Lithuania	1911	1935	Tel Aviv	Jerusalem
Gordon Aharon David	c	Russian Empire	1856	1904	Ein Ganim	Petah Tikva, Rishon LeZion, Degania A
Greenberg Uri Zvi	c	Austria-Hungary	1896	1939	Ramat Gan	
Hadar Kimchy	m	Palestine	1929	-	Ein Harod	
Haim Ben-David	m	Austria	1919	1938	Beisan	
Haim Cohen-Meguri	p	Yemen	1913	1921	Nathanya	
Haim Gouri	c	Palestine	1923	-	Beit Alfa	Tel Aviv, Givat HaSlosha, Kfar Tavor
Haim Hazaz	c	Russian Empire	1898	1931	Jerusalem	
Haim Joseph Zadok	p	Poland	1913	1935	Jerusalem	
Haim-Moshe Shapira	p	Belarus	1902	1925	Jerusalem	
Hammer Zevulun	c, p	Palestine	1936	-	Haifa	Jerusalem
Hamuda	p, m	Palestine	1921	-	Jerusalem	

Continued on next page

Table C1: Leaders characteristics (Continued)

Ish-Shalom						
Hankin Yehoshua	c	Russian Empire	1864	1887	Gdera	Tel Aviv, Rishon LeZion, Rehovot
Hannah Lamdan	p	Yugoslavia	1905	1926	Holon	Tel Aviv
Hannah Szenes	m	Hungary	1921	1939	Nahalal	Sdot Yam, Kiryat Haim
Haviva Reik	m	Czechoslovakia	1914	1939	Maanit	Kibbutz Yagur
Hefer Chaim	c	Poland	1925	1936	Raanana	Tel Aviv
Herzl Berger	p	Russian Empire	1904	1934	Tel Aviv	
Herzl Shafir	m	Palestine	1929	-	Tel Aviv	
Herzog Yitzhak Ha Levi	c	Russian Empire	1888	1936	Jerusalem	
Heshel Frumkin	p	Russian Empire	1896	1920	Deganiya	
Hillel Kook	c, p	Russian Empire	1915	1924	Affule	Jerusalem
Hissin Chaim	c	Russian Empire	1865	1882	Mikve Yisrael	Rishon LeZion, Gdera Tel Aviv
Hulda Gurevich	m	Palestine	1927	-	Athlith	
Idov Cohen	p	Romania	1909	1940	Tel Aviv	
Israel Bar-Yehuda	p	Ukraine	1895	1926	Yagur	
Israel Barzilai	p, m	Poland	1913	1934	Negba	
Israel Kargman	p	Russian Empire	1906	1929	Haifa	Petah Tikva, Magdiel Jerusalem, Kiryat Haim
Israel Rokach	p	Palestine	1886	-	Tel Aviv	
Isser Harel	m	Belarus	1912	1930	Shfayim	Herzliya
Jabr Moade	p, m	Palestine	1919	-	Alumot	
Jacob Fichman	c	Moldova	1881	1919	Rehovot	
Jenia Tversky	p	Russian Empire	1904	1923	Haifa	
Joseph Aharon Almogi	m	Poland	1910	1930	Kfar Saba	Haifa
Joseph H. Brenner	c	Russian Empire	1881	1909	Jerusalem	Hadera, Tel Aviv
Joseph Kremerman	p	Palestine	1926	-	Haifa	
Joseph Shprinzak	p	Russian Empire	1884	1910	Tel Aviv	
Kadish Luz	m	Belarus	1895	1920	Deganiya	
Kalman Kahana	p	Poland	1910	1938	Hafetz Hayim	
Karelitz Avraham Yeshayahu	c	Russian Empire	1878	1933	Bnei Brak	
Katriel Katz	m	Poland	1908	1924	Jerusalem	Ein Harod
Kollek Teddy	c, p	Hungary	1911	1935	Kibbutz Ein Gev	Jerusalem
Kook Avraham Yitzhak	c	Russian Empire	1865	1904	Tel Aviv	Jerusalem
Laskov Haim	c, m	Soviet Russia	1919	1925	Haifa	
Levi Eshkol	p	Ukraine	1895	1913	Deganiya	Petah Tikva, Atarot, Kfar Uriya Rishon LeZion, Degania
Max Buchmann	m	Germany	1896	1924	Tiberias	
Meir Amit	p	Palestine	1921	-	Alonim	Tel Aviv
Meir Argov	p	Ukraine	1905	1927	Petah Tikva	Nes Ziona
Meir Vilner	p	Poland	1918	1938	Jerusalem	
Meir Zorea	p, m	Moldova	1923	1925	Haifa	Tel Aviv, Deganiya, Geniger Kiryat Haim, Qabarra
Menachem Begin	p	Belarus	1913	1943	Tel Aviv	
Menachem Ratzon	p, m	Palestine	1919	-	Petah Tikva	
Menahem Cohen	p	Palestine	1922	-	Jerusalem	
Mordechai Bentov	p	Poland	1900	1920	Mishmar Haemek	Ein Ganim, Nahalel Affule
Mordechai Gazit	m	Turkey	1922	1932	Jerusalem	
Mordechai Maklef	m	Palestine	1920	-	Motza	Pardes Hanna
Mordechai Namir	p	Ukraine	1897	1924	Tel Aviv	
Mordechai Zar	p	Iran	1914	1935	Jerusalem	
Moshe Aram	p	Russian Empire	1896	1924	Tel Aviv	
Moshe Carmel	p	Poland	1911	1924	Naane	Tel Aviv
Moshe Dayan	p	Palestine	1915	-	Nahalal	Kibbutz Hanita
Moshe Kelmer	p	Poland	1901	1921	Jerusalem	

Continued on next page

Table C1: Leaders characteristics (Continued)

Moshe Kol	p	Russian Empire	1911	1932	Kfar Saba	
Moshe Sharett	p, m	Ukraine	1894	1906	Tel Aviv	
Moshe Smilansky	c	Russian Empire	1874	1891	Rehovot	
Moshe Sneh	p	Poland	1909	1940	Tel Aviv	
Moshe Tzadok	m	Poland	1913	1933	Kfar Menahem	Haifa
Nahum Nir	p	Poland	1884	1925	Tel Aviv	
Nahum Shadmi	m	Moldova	1898	1921	Menachemia	Tel Aviv, Bethany Illit Kibbutz Huldah, Kibbutz Gebat, Kefar Vitkin
Naomi Shemer	c	Palestine	1930	-	Kvutzat Kinneret	Tel Aviv
Nathan Alterman	c	Poland	1910	1925	Tel Aviv	
Peretz Goldstein	m	Romania	1923	1940	Afikim	Kinnereth, Kibbutz Maagan
Peretz Naftali	p	Germany	1888	1933	Tel Aviv	
Pessah Bar-Adon	m	Poland	1907	1925	Merhavya (Kibbutz)	
Pinhas Lavon	p	Poland	1904	1929	Hulda	
Pinhas Rosen	p	Germany	1887	1931	Tel Aviv	
Pinhas Sapir	p	Poland	1906	1930	Kfar Saba	
Raanan Weitz	c	Palestine	1913	-	Jerusalem	Rehovot, Moshav Ora
Rachel Tsabari	p	Palestine	1909	-	Tel Aviv	
Rachel Yanait Ben-Zvi	m	Russian Empire	1886	1908	Jerusalem	Tel Adashim
Rahel Bluwstein	c	Russian Empire	1890	1909	Rehovot	Kinneret, Degania, Gdera Jerusalem, Tel Aviv,
Rafael Eitan	p	Palestine	1929	-	Tel Adashim	
Rechav'am Ze'evi	m	Palestine	1926	-	Givath Hashlosha	Jerusalem
Refael Bash	p	Latvia	1913	1938	Kfar Blum	
Reuven Sheri	p	Romania	1903	1925	Kfar Hasidim	Haifa, Kiryat Amal, Kfar Saba, Givat Tivon, Rehovot, Jerusalem
Reuven Shiloah	m	Palestine	1909	-	Jerusalem	
Ruth Haktin	p	Ukraine	1901	1924	Tel Yosef	
Sarah Aaronsohn	m	Palestine	1890	-	Zikhron Yaakov	
Senetta Yoseftal	p	Germany	1912	1938	Givath Hayim (Kibbutz)	Gal Ed
Shabtai Levy	c	Turkey	1876	1894	Petah Tikva	Zichron Yaakov, Haifa
Shalom-Avraham Shaki	p	Yemen	1906	1914	Jerusalem	Tel Aviv
Shaul Avigur	m	Latvia	1899	1911	Kinnereth	
Shaul Biber	m	Palestine	1922	-	Gan Shmuel	Kibbutz Kiryat Inavim, Afula, Givat HaSlosha, Kfar Tavor
Shaul Tchernichovsky	c	Russian Empire	1875	1931	Tel Aviv	Jerusalem
Shimon Avidan	m	Poland	1911	1934	Ayeleth Hashahar	Kibbutz Ein Hashofet
Shimon Mazeh	m	Belarus	1907	1931	Kfar Marmorek	Rehovoth
Shimon Peres	p	Poland	1923	1934	Tel Aviv	Ben Shemen, Kibbutz Geva
Shimshon Unichman	p	Poland	1907	1935	Rosh Pina	
Shlomo Brener	m	Ukraine	1911	-	Haifa	
Shlomo Goren	m	Poland	1917	1925	Harosheth Hagoyim	Kfar Haissidim, Jerusalem
Shlomo Hillel	p	Iraq	1923	1934	Jerusalem	Deganiya, Kibbutz Giva
Shlomo Lahat	m	Germany	1927	1933	Kfar Marmorek	Rehovoth
Shlomo Shamir	m	Ukraine	1915	1925	Jerusalem	
Shlomo-Jacob Gross	p	Hungary	1908	1950	Bnei Brak	
Shmuel Eyal	m	Belarus	1922	1924	Rishon Lezion	
Shneor Zalman Shazar	p	Russian Empire	1889	1924	Tel Aviv	
Stef Wertheimer	c	Germany	1926	1937	Tel Aviv	Tel Litvinsky
Tamar Eshel	m	UK	1920	1923	Haifa	
Tuvia Avramson	m	Serbia	1901	1905	Kinnereth	Rishon Lezion, Tel Hal, Kfar Giladi, Haifa
Tzipora Laskov	p, m	Ukraine	1904	1928		
Tzvi Tzur	m	Soviet Russia	1923	1925	Tel Aviv	
Yaacov Meridor	p	Poland	1913	1932	Petah Tikva	Haifa, Tel Aviv, Rehovoth,

Continued on next page

Table C1: Leaders characteristics (Continued)

						Ramat Gan, Tel Aviv, Magdiel, Ranaana
Yaakov Dori	m	Ukraine	1899	1905	Haifa	
Yaakov Klivnov	p	Belarus	1887	1921	Haifa	
Yaakov Meir	c	Palestine	1856	-	Jerusalem	
Yaakov Nitzani	p	Bulgaria	1900	1935	Tel Aviv	
Yaakov Peri	m	Palestine	1915	-	Nathanya	
Yaakov-Shimshon Shapira	p	Ukraine	1902	1924	Givath Hashlosha	Petah Tikva, Jerusalem Haifa, Tel Aviv
Yechezkel Hen	p	Russian Empire	1908	1925	Tel Aviv	
Yehuda Arazi	m	Poland	1907	1924	Tel Aviv	
Yehuda Leib Maimon	c, p	Russian Empire	1875	1913	Tel Aviv	Jerusalem
Yehudit Simhoni	p	Russian Empire	1902	1921	Nahalal	Kibbutz Tel Yosef
Yerachmiel Assa	p	Soviet Russia	1919	1924	Tel Aviv	Kibbutz Holta
Yeshayahu Foerder	p	Germany	1901	1933	Tel Aviv	
Yeshayahu Gazit	m	Palestine	1925	-	Athlith	al-Lajjun
Yigael Yadin	c, p, m	Palestine	1917	-	Jerusalem	
Yigal Allon	m	Palestine	1918	-	Kfar Tavor	Masha, Kibbutz Ginosar
Yisrael Galili	p	Ukraine	1911	1914	Naane	Tel Aviv
Yisrael Yeshayahu	p	Yemen	1908	1929	Rishon Lezion	Karmia, Tel Aviv
Yitzhak Ben-Aharon	p, m	Austria	1906	1928	Givath Hayim (Kibbutz)	Kibbutz Hashomer Ha'Tsa'ir
Yitzhak Ben-Yaakov	m	Ukraine	1891	1908	Deganiya	Zichron Ya'akov, Haifa, Jerusalem, Menachemiah
Yitzhak Ben-Zvi	p	Ukraine	1884	1907	Tel Aviv	Jerusalem
Yitzhak Coren	p	Russian Empire	1911	1940	Tel Aviv	
Yitzhak Navon	m	Palestine	1921	-	Jerusalem	
Yitzhak Rabin	p	Palestine	1922	-	Tel Aviv	Givat Hashalosha, Ramat Yohanan
Yitzhak Raphael	m	Poland	1914	1935	Jerusalem	
Yitzhak Sadeh	m	Russian Empire	1890	1920	Jerusalem	Nahraim, Athlith, Tel Aviv, Jaffa, Kibbutz Naane
Yitzhak Shamir	p	Poland	1915	1935	Tel Aviv	
Yitzhak Tabankin	p	Belarus	1888	1912	Kinnereth (Moshava)	Merhavia, Kfar Uria, Rehovoth, Ein Harod
Yitzhak-Meir Levin	p	Poland	1893	1940	Jerusalem	
Yizhar Smilansky	p, m	Palestine	1916	-	Kfar Marmorek	Hulda, Tel Aviv, Rehovoth
Yoram Aridor	c, p	Palestine	1933	-	Tel Aviv	
Yosef Burg	p	Germany	1909	1939	Jerusalem	
Yosef Efrati	p	Belarus	1897	1914	Beer Touvya	Petah Tikva, Beer Touvia, Ben Shemem, Geva
Yosef Kushnir	p	Russian Empire	1900	1922	Jerusalem	Haifa
Yosef Sapir	p	Palestine	1902	-	Petah Tikva	
Yosef Serlin	p	Poland	1906	1933	Tel Aviv	
Yosef-Michael Lamm	p	Poland	1899	1939	Tel Aviv	
Yoseph Avidar	m	Poland	1906	1925	Jerusalem	Tel Aviv
Zalman Aran	p	Ukraine	1899	1926	Tel Aviv	Haifa, Kfar Hassidim
Zalman Ben-Yaakov	p	Poland	1897	1935	Tel Aviv	
Zalman Susayeff	p	Latvia	1911	1935	Tel Aviv	
Zeev Shefer	p	Russian Empire	1891	1913	Ayeleth Hashahar	
Zeev Tsur	p	Poland	1911	1931	Rishon Lezion	Sde Nahum
Zerach Warhaftig	p	Russian Empire	1906	1944	Jerusalem	
Zvi Ayalon	m	Russian Empire	1911	1923	Tel Aviv	Haifa
Zvi Dinstein	c, p	Palestine	1926	-	Tel Aviv	
Zvi-Hugo Lehrs	m	Germany	1904	1933	Safad	Jerusalem, Tel Aviv

Notes: In column *Role*, c: cultural leader; m: military leader, p: political leader