State Building in a Diverse Society*

Ruth Carlitz† Ameet Morjaria‡ Joris Mueller§ Philip Osafo-Kwaako¶

First draft: November 2022
This version: August 2024

Diversity can pose fundamental challenges to state building and development. The Tanzanian *Ujamaa* policy — one of post-colonial Africa's largest state-building experiments — addressed these challenges by resettling a diverse population in planned villages, where children received political education. We combine differences in exposure to *Ujamaa* across space and age to identify long-term impacts of the policy. Analysis of contemporary surveys shows persistent, positive effects on national identity and perceived state legitimacy. Our preferred interpretation, supported by evidence that considers alternative hypotheses, is that changes to educational content drive our results. Our findings also point to trade-offs associated with state building: while the policy contributed to establishing the new state as a legitimate central authority, exposure to *Ujamaa* lowered demands for democratic accountability and did not increase generalized inter-ethnic trust.

JEL Codes: I28, J15, N47, P21, Z18
Key words: Diversity, State building, Nation building, Identity, Public education, Africa, Democracy

* We are especially grateful to the editor, Noam Yuchtman, and four anonymous referees for many invaluable comments and suggestions. We thank Sam Asher, Ralph Austen, Tim Besley, Georgy Egorov, Kai Gehring, Leander Heldring, Joel Mokyr, Beatrice Montano, Nathan Nunn, Elias Papaioannou, Daniel Posner, Michael Powell, Nancy Qian, Bryony Reich, Paul Schaudt, Cory Smith, Tuan-Hwee Sng, Christopher Udry, Léonard Wantchekon, and conference and seminar participants at WGAPE, UEA, Governance & Local Development (Gothenburg), Australasian Public Choice (Perth), Barcelona GSE Summer Forum, AMES, AFES, Northwestern, Zurich, Lausanne, Geneva, and Bern. We especially thank Ted Miguel, Tina Green and Mike Shand for sharing data and many conversations. We are particularly grateful for outstanding research assistance by Giacomo Marcolin, as well as Elizabeth Cameron, Mary-Crescent Carter, Riddhimaa Gupta, Mathias Ngowi, Areebah Omer, Sanjith Prabhu, and Tenny Tsang. We thank staff at the Institute for Democracy, Citizenship, and Public Policy in Africa, at the University of Cape Town, and the Herskovits Library of African Studies at Northwestern University, and William Thompson and Ambreen Chaudhri at Kellogg Research Support. Morjaria gratefully acknowledges funding from Northwestern's Global Poverty Research Lab, Ford Motor Company Center for Global Citizenship at Kellogg School of Management and CIFAR's Azrieli Global Scholars Program. Mueller gratefully acknowledges funding from Ministry of Education AcRF Tier 1 grant FY2023-FRC2-006.

† University of Amsterdam: r.d.carlitz@uva.nl

‡ Kellogg School of Management, CEPR and NBER: a.morjaria@kellogg.northwestern.edu

§ National University of Singapore: jorismueller@nus.edu.sg

¶ Africa Center for Economic Transformation: osafo@alum.mit.edu

1. Introduction

Diversity can pose fundamental challenges to state building and development, including the under-provision of public goods due to heterogeneous preferences (Alesina et al., 1999), inter-group conflict driven by inequality and competition for resources (Ray and Esteban, 2017), and difficulties with communication that hinder the establishment of social trust (Wimmer, 2018). In more diverse polities, citizens may not feel that public policies represent their preferences sufficiently (Alesina and Spolaore, 1997), posing a challenge to state legitimacy (Holsti, 1996). Such a lack of perceived state legitimacy is in turn associated with weak state capacity, poor economic performance, and even violence (Besley and Persson, 2011).

To overcome these challenges, political leaders throughout history have worked to build new *states* through *nation*-building endeavors that include the promotion of common values, identity, and language (often through public education) and population resettlement (Tilly and Ardant, 1975; Smith, 1986; Hobsbawm, 1992).[1] However, the implementation of such measures requires state capacity and reach, which may be hindered by diversity in the first place (Miguel and Gugerty, 2005). In addition, political leaders may encounter resistance from elites and embedded group-specific social norms. Leaders from early modern Europe to the post-colonial era have therefore relied on bundles of "homogenizing, territorializing, and mobilizing" activities (Smith, 1986) to address multiple challenges simultaneously.[2] It is an empirical question what the long-run results of these activities are.

This question is difficult to answer since nation-building activities typically unfold over long time spans and affect entire populations simultaneously. While scholars have identified causal effects of specific elements of nation-building policies, states generally implement such policies as bundles of activities. Moreover, interactions between different components of a given bundle may lead to combined effects distinct from the sum of their parts. This paper aims to identify such combined effects.

We study a setting that provides sharp variation in exposure to a bundled state-building effort. Specifically, we investigate the long-run effects of one of post-colonial Africa's most ambitious state-building experiments — the Tanzanian *Ujamaa*[3] policy — on the development of national identity in a highly diverse society. We also examine whether, in building the Tanzanian *nation*, *Ujamaa* strengthened the nascent Tanzanian *state*, looking at how *Ujamaa*

[1] We understand state building as the construction of a state apparatus that can establish a monopoly on the legitimate use of violence, in order to protect property rights, collect taxes, and provide public goods in a given territory (Tilly and Ardant, 1975). Nation building captures the formation of a national identity with which citizens feel a sufficient sense of emotional commonality that they wish to remain together. Nation building may facilitate state building (Alesina et al., 2021).

[2] Examples include efforts in the 19th-century to turn "peasants into Frenchmen" (Weber, 1976) and "make Italians" following that country's unification (Duggan, 2008), as well as the 20th-century construction of national identity in China (Wimmer, 2018) and the Soviet Union (Martin, 2001).

[3] *Ujamaa* roughly translates as "familyhood" (Sitari, 1983).

affected citizens' attitudes towards and engagement with the state and state institutions in the long run. We digitize historical administrative data and combine it with contemporary survey data to operationalize our empirical strategy.

While the political economy of state building in diverse societies is globally relevant, it is particularly pertinent in post-colonial Africa. Newly independent African states inherited artificial borders that contained a multitude of ethnic groups with little history of centralized governance or strong shared identity (Alesina et al., 2011).[4] The salience of ethnic differences had been "reinforced [and] exaggerated" (Mamdani, 2003, p. 139) to serve the interests of the colonizers. As a result, post-independence leaders in Africa faced a number of challenges to state building: fragmented populations scattered over inhospitable territories beyond the state's reach (Herbst, 2014), a mix of potentially rival ethnic groups within arbitrary borders (Michalopoulos and Papaioannou, 2016), and no workable social contract between the state and population (Scott, 1998). These challenges made it difficult for many new African states to build state capacity and develop common values around which to unite their diverse populations.

The *Ujamaa* policy, implemented from 1970 to 1981, addressed these challenges simultaneously by bringing the Tanzanian population into the state's reach and forging a national identity through the public education system. The key tool for expanding the state's reach was the so-called 'villagization' program, which resettled much of the multi-ethnic rural population in planned villages under state administration. Villagization was remarkably successful, with over 80 percent of the country's rural population living in planned villages within just a few years. As Scott (1998) explains, "the purpose of forced settlement is always disorientation and then reorientation" (p. 235). In the Tanzanian case, "reorientation" was facilitated by abolishing traditional authorities and replacing them with state bureaucrats and democratically elected village governments (Sitari, 1983).[5] Critically, *Ujamaa* also put education in the planned villages into the hands of the new state, which concurrently revamped the curriculum to reflect its political goals. The *Ujamaa* policy thereby allowed the state "to capture the peasantry" both geographically and ideologically (Hydén, 1980).

Exposure to *Ujamaa* varied both across space and across age cohorts. Villagization intensity varied across districts as implementation was largely left to local officials, who enforced resettlement with different degrees of conviction. Sharp and plausibly exogenous variation can be seen across cohorts given the policy's inclusion of a targeted education reform. Only those young enough to enter school after the policy's introduction were exposed to the entire state-building bundle, including the revamped curriculum in the planned villages.

[4] Afrobarometer data, from rounds 3 and 4, reveals that over 50% of the African population identifies more with their ethnic group than with their nation.

[5] Other elements of the *Ujamaa* policy included the abolition of individual titles to property, collectivization of agricultural production, and the nationalization of certain enterprises.

We are thus able to study impacts of *Ujamaa* using a difference-in-differences cohort design in the spirit of Duflo (2001). Our empirical specification interacts time-invariant local villagization intensity with birth cohort indicators. We neither assume that levels of villagization intensity varied randomly across space, nor that there were no underlying differences between different cohorts in the absence of villagization. We control for systematic differences between districts and cohorts with district and region-specific cohort fixed effects. Only the interaction between the two sources of variation constitutes exogenous variation in citizens' exposure to the *Ujamaa* policy under a parallel trends assumption. The sharp timing and the short-lived nature of the policy allow us to compare differences in outcomes for cohorts that were of primary school age during the policy with cohorts that were slightly too young or too old to be affected by the bundle of activities, for more and less intensely villagized areas.[6]

We find large, significant, and persistent positive effects of *Ujamaa* on national identity. In surveys conducted around two decades after the policy ended, a one standard deviation (SD) increase in villagization leads members of the primary school-age cohort to identify 0.23 SD more strongly with the Tanzanian nation rather than with their ethnic group, compared to the control cohort.[7]

The result withstands a battery of robustness checks. We find no differences in the trends across cohorts of primary school age before and after the policy period for more and less intensively villagized districts, lending support to the parallel trends assumption. In addition, our result is robust to controlling for pre-policy district characteristics interacted with cohort fixed effects. We also conduct a bounding exercise which shows that selective migration is unlikely to explain a large part of the estimated effect. Finally, we control for respondents' beliefs about surveyors to reduce concerns related to social desirability bias.

We interpret our main coefficient as the differential effect of the bundled treatment compared to the effect of villagization alone, i.e., holding constant the average effect of villagization in the overall population.[8] While villagization alone may in theory have affected national identity across all cohorts, only those of primary schooling age during *Ujamaa* were exposed to both villagization and the education reform. The fixed effects in our specification control for the variation in the intensity of villagization across space as well as overall

[6] Our empirical strategy exploits differences across districts in villagization intensity at the height of the *Ujamaa* period but not differences across districts in the timing of implementation, as this is not clearly documented. The variation in timing stems from variation in respondents' birth cohorts.

[7] The difference between national identity of the treated and the control cohort, for a district with complete villagization compared to a district without any villagization, corresponds to the difference between a respondent "feeling more Tanzanian" and "feeling more a member of their ethnic group [than Tanzanian]." We further find that men in treated cohorts in districts villagized by 1 SD more are 6.1 percentage-points more likely to marry women from ethnic groups different from their own.

[8] The effect of villagization alone appears to be small. Average levels of national identity are no higher in more intensively villagized districts for cohorts outside of primary schooling age during the *Ujamaa* policy.

differences across cohorts. Our results thus imply that villagization served to make the education reform a more effective means of bolstering national identity. We understand this as resulting from the disruption of existing norms and hierarchies that villagization engendered, including resettlement and the abolishment of traditional village leaders in favor of bureaucrats (Feierman, 1990). This speaks to the bundled nature of the policy, highlighting the importance of possibly interacting activities. Understanding how the pieces of the bundle in isolation affected national identity is not the focus of this paper.

Nevertheless, we examine potential channels driving our results. Our preferred explanation is that *Ujamaa* primarily influenced national identity through political education provided in planned villages. We also explore evidence for other cohort-specific factors that may have interacted with villagization. We find empirical support for our preferred explanation but little evidence for alternatives. In particular, we find that exposure to *Ujamaa* increased national identity only among those who attended primary school when the policy was in effect. Furthermore, we provide evidence that the impact arises specifically from changes to the public school curriculum, rather than from other factors coinciding with primary school attendance, such as intergroup contact or increased human capital.

Moreover, we find a large effect of *Ujamaa* on the cohort that was of primary school age during the policy period, but little effect on the cohorts of primary school age after the policy ended. This implies that changes to the content of education rather than a general increase in its supply explain our results.

Having established that *Ujamaa* helped bolster national identity, we next examine whether the policy also served to foster state legitimacy in the long run. This is an important question given that state builders have used such policies throughout history to attempt to establish authority. Moreover, *Ujamaa*'s effect on state legitimacy is *ex-ante* unclear. Such a policy could either enhance state legitimacy by increasing the population's identification with the nation and the state, or provoke backlash, undermining state legitimacy.

We find that cohorts exposed to *Ujamaa* are more likely to respect state authority and approve of one-party rule, and to have higher trust in government institutions such as state media. Interestingly, we find little effect on generalized inter-ethnic trust. These findings suggest that *Ujamaa* primarily contributed to the new Tanzanian state's ability to consolidate power. However, the policy also reduced demands for democratic accountability.

Tanzania's experience with state-building offers relevant insights for a wide range of countries which face cleavages based on history, geography, and culture (see e.g., Alesina and La Ferrara, 2005; Ray and Esteban, 2017). Such cleavages may contribute to why these countries – many of which have a history of colonialism – struggle to sustain strong states. Our study finds that bundled nation-building policies can help overcome such challenging

conditions, but can come at a cost.[9]

We build on a rich theoretical and historical literature on the processes of nation and state building (e.g., Tilly and Ardant, 1975; Smith, 1986; Anderson, 1991). A more recent economics literature has provided empirical evidence in support of these ideas and theories – primarily identifying causal effects of specific aspects of broader nation-building policies (Rohner and Zhuravskaya, 2023). For example, Bazzi et al. (2019) study a population resettlement program in Indonesia to identify effects of intergroup contact on national integration.[10] Economists studying state building have also turned their attention to Africa. Here, our work is related to Depetris-Chauvin et al. (2020), who investigate the short-term impact of national football teams' victories on national identity in Sub-Saharan Africa. Blouin and Mukand (2019) examine how propaganda broadcast over radio helped to change inter-ethnic attitudes in post-genocide Rwanda.[11] Okunogbe (2024) studies inter-group contact and national integration in the context of national youth service in Nigeria.

Our contribution to the literature on state building is twofold. First, we study a bundle of measures aimed at bolstering national identity and state legitimacy. This is important because policies in practice typically manifest as bundles; as noted above, the effects of specific measures may be very different in isolation compared to when combined with other measures. For example, when enacted in isolation, policies mandating the forced coexistence of diverse ethnic groups have led to backlash and even spurred inter-group conflict (Dippel, 2014).[12] The lack of widespread backlash to *Ujamaa* speaks to the bundled nature of the policy, highlighting the importance of interacting activities. Second, we study attitudes towards the state in addition to national identity, finding that effective state building may come at the cost of making citizens comply with or even prefer non-democratic governance. These findings provide new insights for the emerging literature on nation building, which has not extensively examined potential trade-offs vis-à-vis acquiescence to authoritarianism.

This paper also speaks to scholarship on the role of education in state building, which philosophers and politicians have recognized dating back to Plato's *Republic*. Economists

[9] Our findings also have relevance for states without a recent history of colonialism by Western powers. For example, China and Russia have pursued policies to establish strong national identities and state authority in a manner that runs counter to democratization and fostering societal trust (Bahry et al., 2005; Wu and Shi, 2020).

[10] As in newly independent Tanzania, the Indonesian government was concurrently engaged in a number of activities oriented toward nation building. These interventions included amongst others resettlement, school construction, and nationalization of state media.

[11] This was part of a bundled policy attempt to reconcile inter-ethnic relations, which included setting up a National Unity and Reconciliation Commission; national radio; revising the constitution to criminalize divisionism; sites of remembrance to promote collective memory; new national symbols, administrative restructuring and renaming places; rewriting history textbooks; and introducing social programs.

[12] Two notable examples in Africa are post-independence governments of Mozambique and Ethiopia, which both 'villagized' 2 and 12 million people respectively (Lorgen, 2000). In both cases, forced coexistence was followed by the downfall of the ruling political party and sustained conflict.

have also begun studying how education can contribute to national identity and ideology (Cantoni and Yuchtman, 2013) – in a wide range of settings and with mixed outcomes (Cantoni et al., 2017; Clots-Figueras and Masella, 2013; Cinnirella and Schueler, 2018; Bandiera et al., 2019; Fouka, 2020; Blanc and Kubo, 2021; Bazzi et al., 2023). Much of this work is grounded in the ideas put forth by psychologists and sociologists that adolescence comprises particularly "impressionable years", which serve as a formative phase for establishing cultural orientations (Krosnick and Alwin, 1989) and identity (Erikson, 1968).

We contribute to the literature on education and state building in several ways. First, this literature has typically – and by design – isolated changes to education from various other institutional changes that go along with state-building efforts. In contrast, we study an education reform in conjunction with other institutional changes that disrupted existing social structures and hierarchies. Second, we study the effect of education on nation building in a particularly pertinent and understudied context. Much of the existing work in this area has examined education in contexts where the government could tap into an existing shared history or culture (Hobsbawm, 1992; Gellner, 2006). Post-colonial Africa in contrast was characterized by high diversity and little "sharedness" due to artificial borders and the colonial experience, making it an unlikely place for such an approach to nation building.

Our paper bridges two well-established literatures in the social sciences that have previously been largely separate – the economics literature on nation building and education (Alesina et al., 2021) and the anthropological literature on state building (Scott, 1998).[13] We also contribute to literature that traces the impact of pre- and post-colonial institutions on African development. Much of this work examines the negative effects of artificial colonial borders (Easterly and Levine, 1997; Alesina et al., 2011; Michalopoulos and Papaioannou, 2016), particularly the resulting ethno-linguistic fractionalization. Less attention has been given to national institutions that aim at overcoming the problems resulting from colonization (Michalopoulos and Papaioannou, 2013; Burgess et al., 2015). In this context, our paper is one of the first to empirically examine the state-building consequences of *Ujamaa*.

Scholars have predominantly examined *Ujamaa*'s negative economic impacts (Hydén, 1980; Putterman, 1986; Collier et al., 1986; Scott, 1998), with limited causal inference efforts. A notable exception is Osafo-Kwaako's (2012) PhD thesis, which provides an important point of departure for the present study.[14] The thesis examines the impacts of living in previously villagized areas on various social and economic outcomes: it identifies positive

[13] Other contemporary work on nation building examines how religious pilgrimage (Clingingsmith et al., 2009); war, occupation and repression (Dehdari and Gehring, 2022; Abramenko et al., 2024); foreign enemies (Dell and Querubin, 2018; Gehring, 2022); national leaders (Assouad, 2021); political boundaries (Bluhm et al., 2021); and representation and redistributive institutions (Giuliano et al., 2023) shape national identity.

[14] Another study employing a causal inference framework to examine *Ujamaa* is the PhD thesis by Jarotschkin (2018). It utilizes a cross-sectional instrumental variables strategy to investigate the policy's effects on economic development and inter-ethnic trust.

effects on schooling and public goods provision, alongside negative effects on consumption and perceptions of corruption. In contrast to this paper, the thesis finds a positive impact on ethnic (vs. national) identity and a negative impact on support for one-party rule. These contrasting findings are a consequence of the distinct research questions asked by the two studies, which lead to different cross-cohort comparisons and treatment construction: we focus on the effects of *Ujamaa* on state building and compare birth cohorts eligible for political education during *Ujamaa* with older cohorts. We thus study the effects of the *bundle* of villagization interacted with political education, whereas Osafo-Kwaako (2012) studies *Ujamaa* as an exemplar of state-led development planning. The thesis compares cohorts exposed to villagization (but not to political education) or cohorts exposed to political education (but not under villagization) with older cohorts, but not the cohorts exposed to the bundle. This approach captures *Ujamaa*'s legacy and necessarily excludes this paper's treated cohorts (those of the age to receive political education during *Ujamaa*). Given our distinct focus on state building, we also examine different outcomes, including inter-ethnic marriage, respect for state authority, democratic accountability, and trust in the media. Lastly, Miguel (2004)'s influential article, which examines how nation building in Tanzania and neighboring Kenya has affected inter-ethnic cooperation (or the lack thereof) and public goods provision, is also closely related to our work. We complement this work by adding evidence on changed attitudes towards the state and democratic preferences.[15]

This paper proceeds in seven sections. Section 2 describes the *Ujamaa* policy and contextual background. Section 3 presents the data we use to measure exposure to the *Ujamaa* policy and its consequences. Section 4 outlines our empirical strategy to address challenges to identification. Section 5 presents our main result on national identity, conducts several robustness checks, and considers alternative channels. Section 6 discusses implications of our findings on views of the state and the legitimacy of state authority. Finally, Section 7 offers concluding thoughts.

2. Background

Ujamaa comprised a series of reforms in post-independence Tanzania. These included the institutionalization of one-party rule and efforts to promote equality across all spheres of society, such as the nationalization of banks and large industrial enterprises and state price control (Ergas, 1980). Most important for this study – and arguably the most wide-reaching

[15] Miguel (2004) makes a cross-sectional comparison of two nearby districts in Tanzania and Kenya which are separated by a national border but are similar in terms of local ethnic diversity. The study finds a negative relationship between ethnic diversity and public goods provision in Kenya, but a positive relationship in Tanzania. This is explained by Tanzania having implemented effective nation-building policies to ameliorate ethnic divisions, whereas Kenya did not. Focusing on Tanzania, our study provides direct evidence supporting this explanation – further looking at *all* districts on the mainland.

reforms – were the resettling of the rural population into planned villages and harnessing public education as a nation-building tool. These activities aimed at organizing the country around the new state (Nyerere, 1969a,b). The villagization program – implemented primarily between 1970-1981 – mandated the rural population to live in state-administered villages (Scott, 1998). Moreover, in those villages, primary school-age children were exposed to a new political education curriculum centered around building a national identity and establishing the Tanzanian state as a legitimate authority (Nyerere, 1982). We now discuss key elements of the policy. Appendix Table A.1 summarizes the timing of key events.

When Tanzania gained independence from Britain in 1961, the country's leader Julius K. Nyerere encountered a challenge common to many post-colonial states: how to organize diverse people, spread across a large territory, around a common mode of governance. Nyerere's challenge was particularly acute: the country's population comprised over 120 ethnic groups with different languages or dialects (Omari, 1995), making Tanzania one of the world's most diverse countries. Furthermore, much of the country was sparsely populated. As of the 1967 Census, 12 million inhabitants were spread over nearly one million square kilometers, with nearly 95 percent of the population residing in rural areas.

Nyerere developed a multi-faceted agenda to unite his country's diverse and geographically dispersed population, embarking on what Scott (1998) describes as one of the most ambitious nation-building programs in post-colonial Africa. In 1963, Nyerere consolidated political power and declared the governing Tanganyika African National Union (TANU) as the sole legal party, a measure that was understood to reduce societal divisions (Tripp, 1999). Nyerere subsequently expounded his ideas for the Tanzanian state in his 1967 landmark speech, the *Arusha Declaration*. The speech presented his vision of *Ujamaa*, which included the central role of villages for state building. The speech also outlined various measures aimed at reforming the economy in line with socialist principles, such as economic self-reliance, nationalizing commercial farms and businesses, and creating co-operatives rather than individual family farms to avoid class formation.

At the time, the majority of the Tanzanian population resided in scattered hamlets in the rural countryside. The *villagization* initiative was proposed to resettle the rural majority into planned villages. The new *Ujamaa* villages were also intended to promote communal farming with modern agricultural methods, which would generate surplus income to finance social infrastructure. Various inducements, such as the provision of schools, clean water and dispensaries, were used to promote movement to the planned villages.

At the end of 1969, the government began mandatory villagization, though no guidelines for implementation were provided apart from declaring 1976 the target year for full rural villagization.[16] Implementation was largely left to district officials, who pursued the endeavor

[16] Presidential Circular No. 1 (Mung'ong'o 1995:80-1 in Kikula, 1997).

with different degrees of conviction (McHenry, 1979). In 1975, the *Villages and Ujamaa Villages Act* was promulgated, mandating registration of villages with 250 or more households as legal entities (Bryceson, 1982). The Act also ushered in various governance reforms at village level, including the establishment of new decision-making bodies. As a result, "The traditional extended family or kinship group with collective responsibility for its members was replaced... by a self-governing *ujamaa* village divided into ten-house cells, and the traditional chiefs and elders by democratically elected village administration" (Sitari, 1983, p. 2). The Act thus served to further one of Nyerere's key aims since taking power after independence – namely, removing traditional authorities from power (Tripp, 1999). The 1975 Act also abolished existing primary cooperative societies, officially replacing them with the village governments (Hydén, 1980).

As a resettlement initiative, the villagization program was successful, with over 80 percent of the country's rural population living in registered villages by the end of 1976 (McHenry, 1979). Appendix Figure A.1 depicts the rapid growth of the population in registered villages over the period after independence. As Scott (1998) explains, the speed with which villagization was carried out was intended to "rip the peasantry from their traditional surroundings and networks" (p. 235). Remarkably, in most cases this resettlement was not met with violent resistance (Ergas, 1980). Moreover, the average distance people moved was rather small – usually within eight kilometers (Sitari, 1983).

Ujamaa implementation varied across districts in terms of scope and intensity due to largely idiosyncratic factors. This reflects the program's being implemented in a decentralized manner, with the precise strategy left to local officials to formulate (McHenry, 1979). Other drivers of villagization intensity included the provision of drought or flood relief (Lorgen, 2000; Hydén, 1980). Scholars have also noted that certain regions saw very little impact of *Ujamaa* given that farmers were already living in populous villages and that cash crops – which provided vital state revenue – were produced in large quantities in these areas (Scott, 1998; Ergas, 1980; Sitari, 1983). We discuss potential correlates with villagization further in Section 3 (Data) and Section 5 (Results).

Concurrently with villagization, the state centralized and consolidated the education sector. It was clear to Nyerere that to create a strong national identity, the education system had to be part of the process (Miguel, 2004). As the Minister of Education Solomon Eliufoo put it in 1968, education was "a cementing matrix of *Ujamaa* and progress" (Peeples, 2018, p. 46).

The *Ujamaa* education reforms were outlined in Nyerere's second key declaration of 1967, the *Education for Self-Reliance* (ESR) paper. ESR consisted of two key elements: bringing the education sector into the hands of the state and changing the school curriculum to help foster nation building. To execute the ESR, the government nationalized all schools with the Education Act of 1969. While only 7% of the population had completed some public

education by 1967, data from the 1978 census showed almost 90% enrollment (Government of Tanzania, 1984).[17] The revised curriculum was likely more effectively implemented in more villagized areas during the *Ujamaa* period.

The Ministerial Circular of 1968 introduced Political Education (known in Swahili as *Elimu ya Siasa*) as a new subject to the primary school curriculum for students enrolled in Standards IV-VII (equivalent to 4^{th} grade onwards). The introduction of Political Education meant new syllabi and new textbooks, as well as new approaches to teaching social studies, history, geography, and civics (Mbunda, 1982). However, the directive to overhaul the curriculum was not necessarily accompanied by additional resources. As Komba (1996, pg. 108) notes, "teachers were advised to use party documents, namely the Arusha Declaration, *Ujamaa Vijijini* (meaning, rural socialism), and Education for Self Reliance."

Political Education served primarily as a tool for ideological indoctrination, emphasizing the importance of the nation and the state over tribal identities (Komba, 1996). In addition, Education Circular No. 2 of 1967 mandated Swahili as the national language of instruction in all public schools by November 1969. Analysis of educational materials during the *Ujamaa* period reveals emphasis on a state-centric version of national identity. As Komba (1996, p. 111-112) notes, "the dominant theme [of the Political Education syllabus] was, obviously, nationalism," and, "the general tendency was in the direction of political indoctrination rather than genuine Political Education." Appendix Figure A.2 enumerates the contents of a typical Political Education textbook, highlighting the themes that speak to nation and state building. Appendix Figure A.3 depicts excerpts from a Swahili textbook of the period, illustrating how nationalist themes were found in all subjects.[18]

Economic failures associated with the villagization campaign and attempts to organize the country's economy along socialist lines (Hydén, 1980; Collier, 1988) ultimately led to the end of *Ujamaa*. This was codified with the repeal of the *Villages and* Ujamaa *Villages Act* in 1982. However, the end of *Ujamaa* was arguably hastened by key external factors. These included the surprise invasion by Uganda under Idi Amin in 1978 (Roberts, 2014) and the introduction of an IMF economic liberalization program (Kaiser, 1996). Tanzania was a victim of the global debt crisis, which began when the Federal Reserve increased U.S. interest rates from 10 percent to over 20 percent between 1979 and 1982. This policy would have drastic consequences for Tanzania and other countries in the region, as debt servicing

[17] Data on historical school construction is unavailable.

[18] The new political education curriculum was delivered alongside efforts to spread government propaganda through the media. Indeed, Nyerere's 1967 *Ujamaa* manifesto listed the news media as a key means of production and exchange to facilitate his vision (Sturmer, 1998). In 1965, the Tanganyika Broadcasting Corporation was renamed Radio Tanzania Dar es Salaam (RTD) and inaugurated as a department of the Ministry of Information, Broadcasting and Tourism. In 1970, Nyerere nationalized the country's most influential newspapers, *The Standard* and its sister *Sunday News*, which up until then had been under foreign ownership. That said, the media channel had relatively limited reach compared to the new school curriculum, not the least due to unavailability of newspapers in most of the countryside.

costs crowded out social spending. Then Federal Reserve Chair Paul Volcker later noted that "Africa was not even on my radar screen" (Brautigam, 2023). The end of *Ujamaa* ultimately led to a shift in the country's political and economic orientation away from socialism and ideological indoctrination, which has been sustained by subsequent governments (Costello, 1996). This shift also led to the abandonment of the Political Education curriculum, and ultimately, the official reintroduction of Civics in its place in 1992 (Komba, 1996).

As we show in what follows, the *Ujamaa* experiment – though spanning little more than a decade – had a profound and lasting impact on the hearts and minds of Tanzanian citizens, particularly those exposed to public education in planned villages during the *Ujamaa* period.

3. Data

Our sample includes individuals responding to contemporary surveys, from which we obtain outcomes and birth dates to capture variation in their exposure to the *Ujamaa* policy over time. Using information on respondents' place of residence, we link this sample to historical census data to capture variation in their exposure to villagization over space. Table 1 provides descriptive statistics. Appendix Tables A.2 and A.3 provide an overview and description of the variables used in the paper.[19]

3.1. Historical District-level Data

We use newly digitized data from the 1978 population census (Bureau of Statistics, Ministry of Planning and Economic Affairs, 1981) to measure the historical extent of villagization across space. The historical intensity in an individual's district of residence proxies for their exposure to villagization. While we have also obtained data on villagization from the historical population censuses at finer geographical units, it cannot be linked to outcomes from contemporary surveys, which do not typically identify respondents at levels lower than the district.[20] Appendix Figure A.4 shows an excerpt of the 1978 census data. We measure villagization as the share of a district's rural population living in registered government villages in 1978:

[19] Due to the rural nature of the policy, we exclude the capital Dar es Salaam (Mzizima) as well as the semi-autonomous islands (Zanzibar, Pemba, Mafia) from our sample. The islands' governments have separate authority over a number of government functions, including education. Moreover, we exclude from all analyses the districts Chunya and Mufindi, as data on two of our pre-policy controls from Table 2 are missing for these districts. These controls are primary school enrollment in 1967, missing for the Mufindi district, and hospital beds in 1967, missing for the Chunya district.

[20] We do not observe respondents' district of birth or district of residence during the *Ujamaa* policy. We discuss in further detail below how our using the current district of residence affects the interpretation of our estimates and potential threats to identification, including selective migration.

$$V_d = \frac{P_{d,1978}^{registered}}{P_{d,1978}^{rural}}$$

where $P_{d,1978}^{registered}$ denotes the number of individuals living in registered villages in district d and $P_{d,1978}^{rural}$ denotes total rural district population in district d in 1978. We aggregate districts to their 1967 boundaries since this is the unit of variation of our pre-policy controls.

As shown in Table 1, Panel A the villagization measure has a mean of 0.95, and significant dispersion with a minimum of 0.52 and a maximum of 1. The average extent of villagization is relatively high because some geographic zones of the country were completely villagized. However, there is considerable within-zone variation in other areas, which we exploit in our empirical analysis. Appendix Figure A.5 illustrates the variation.[21]

Variation in the extent of villagization across space can largely be explained by the fact that its implementation was left to district officials (a further reason for our focus on the district as a unit of analysis), who enforced the policy with different degrees of conviction (McHenry, 1979). This variation may not be random across space, nor does it need to be for our empirical strategy to be valid, as we explain in Section 4. Nevertheless, in our analysis below we control for the most important potential correlates of villagization intensity, as identified in historical accounts. These include: pre-*Ujamaa* primary school enrollment rate, distance to capital (to capture central government influence), district revenues (to capture local government capacity), ethnolinguistic fractionalization, geographic characteristics, availability of public health infrastructure, weather shocks, pre-policy feelings of national identity, the degree of urbanization, population density, GDP and agricultural production (to capture economic development), distance to Uganda (to capture potential exposure to the 1978-79 Uganda-Tanzania conflict), and the presence of colonial missions and roads (Ergas, 1980; Hydén, 1980). These baseline characteristics variables are mostly digitized from historical district-level government statistics (Jensen, 1968) based on the 1967 population census or from the 1967 population census itself (Central Statistical Bureau, Ministry of Economic Affairs and Development Planning, 1968), and complemented by data from various other sources. We discuss these correlates in Appendix A.

3.2. Individual-level Survey Data

Our primary data on long-run outcomes and other individual-level characteristics is from the geo-coded Afrobarometer public opinion survey Rounds 3 (2005) and 4 (2008) (Afro-

[21] Tanzania has two main subnational administrative units: the region and the district. A zone is a larger subnational geographic area that is not an official administrative unit but is commonly used by organizations such as the Demographic and Health Surveys (DHS) and the Tanzania Ministry of Health. A zone contains three regions or six districts on average.

barometer, 2017; BenYishay et al., 2017).[22] The Afrobarometer surveys are widely used, nationally representative surveys conducted by a pan-African research institution unaffiliated with any national government. The survey rounds we use comprise a sample of 1,797 individual respondents, born between 1948 and 1987, from across Tanzania. We focus on respondents born between 1948-1987, which ensures that the individuals in the sample would have been able to complete their primary education in the post-independence period. We use information on individuals' districts as reported in the survey to match them to our district-level data. Our focal dependent variable captures national identity and is based on the following question:

"Let us suppose that you had to choose between being a Tanzanian and being a [respondent's previously reported ethnic group]. Which of the following statements best expresses your feelings?"

Respondents could report that they feel *only/more/equally/less/not at all Tanzanian* as compared to their ethnic group. We code our baseline outcome variable on a 0 to 1 scale with quarterly increments, where 0 indicates that the respondent identifies only with their ethnic group and 1 indicates that they identify exclusively with Tanzania.[23] We interpret a higher score as reflecting a stronger national identity. This measure of national identity is standard in the literature (see, e.g., Depetris-Chauvin et al., 2020).

We use several other variables from the Afrobarometer survey data to investigate channels and measure outcomes, such as respondents' views of the state and trust in the media.[24] We describe these variables as we introduce them for the empirical analysis in Sections 5 and 6. We also use respondents' birth years from the Afrobarometer survey to measure their temporal exposure to the *Ujamaa* policy. Finally, we construct a dummy variable for whether the respondent has completed primary schooling, which we use in complementary analyses.

Table 1, Panel B shows individual-level descriptive statistics for the baseline sample used in our analysis. The mean measure of national identity is 0.89. National identity is high on average in Tanzania compared to the rest of Africa, presumably in part due to the *Ujamaa* policy. The average birth year of respondents in our sample is 1962. Most of the respondents (88%) completed primary school.

[22] Our choice of these rounds reflects data availability (key questions were incomparable in their wording in Rounds 1 and 2) and time since the villagization policy ended.

[23] Moreover, we code the variable as 1 for respondents who do not identify with any ethnic group according to another survey question. The question in the main text quoted above is based on the Round 4 codebook. The text for Round 3, which we code up analogously, is as follows: *Let us suppose that you had to choose between being a Tanzanian and being a [respondent's previously reported ethnic group]. Which of these two groups do you feel most strongly attached to? I feel only/more/equally [respondent's group] than Tanzanian or more/only Tanzanian.*

[24] The sample size for the analysis of some of these questions is smaller because they were only asked in one rather than both Afrobarometer rounds used in this paper.

3.3. Other Data

We incorporate data from various other sources for additional analyses and robustness checks. These include Round 1 of the Tanzania National Panel Survey (TNPS) in 2008/2009 (National Bureau of Statistics, 2010), which is part of the World Bank's Living Standard Measurement Surveys and includes several thousand respondents from across Tanzania. We use this dataset in robustness checks related to migration.[25] We also use data on individuals' occupations and household consumption from the TNPS to investigate channels. The majority of the respondents work in agriculture (82%).

We use data from the Demographic and Health Surveys (DHS) in 1991 and 1996 (Ngallaba et al., 1993; Bureau of Statistics/Tanzania and Macro International, 1997) to study intermarriage. These two survey waves report married couples' ethnic affiliations (later waves do not report ethnicity). We code a dummy that equals 1 if a respondent shares his or her spouse's ethnic group, and 0 otherwise. In our sample, between 66% and 70% of marriages occur within the same ethnic group, depending on whether we select the sample based on the husband's or wife's cohort.

Finally, we use data from the IPUMS sample of the 1988 population census (Minnesota Population Center, 2015) to study educational attainment. We use data on electoral outcomes from the National Electoral Commission in Tanzania (Carlitz, 2017), which we describe in Section 5.

4. Empirical Strategy

This section outlines our empirical strategy to estimate effects of the *Ujamaa* policy, which included villagization and a public education reform, on national identity and attitudes towards the state in the long run. The ideal experiment would randomly assign *Ujamaa* to some communities or individuals but not to others, and then compare their outcomes. However, *Ujamaa*, like most policy reforms throughout history, was not carried out in such a manner and thus there may be joint determinants of villagization, exposure to education, and outcomes of interest. We use a difference-in-differences specification to address such confounders. The first difference is spatial and comes from varying intensity in villagization across individuals' home districts. The second difference is temporal and comes from plausibly exogenous variation in the exposure of age cohorts induced by the timing of the *Ujamaa* policy (and in particular the curriculum reform). We discuss our identifying assumptions and potential threats to identification after first introducing our specification.

Our empirical strategy relies on the fact that the date of birth and villagization intensity in an individual's district jointly determine their exposure to *Ujamaa*. Individuals born in

[25] The Afrobarometer data do not contain information about migration history or birth district.

1959 or earlier were older than the earliest official political education age when the *Ujamaa* policy took effect in 1970. They should thus not be affected by the full "treatment." In contrast, individuals born between 1960 and 1971 were young enough to be exposed to both villagization and the government's new political education curriculum, which was taught from 4^{th} grade. They were also old enough to attend 4^{th} grade before the villagization period officially ended in 1982. In the main specification, our control cohort was thus born in 1948–1959, and our treated cohort was born in 1960–1971. In complementary analyses, we also consider other cohorts. The two cohorts in our baseline sample consist of 849 individuals in 52 districts.

4.1. Identifying Effects of Ujamaa

For our baseline, we estimate the following specification:

$$y_{idzts} = \beta(V_{dz} \cdot treatedcohort_t) + (\mathbf{X'}_{dz} \cdot treatedcohort_t)\mathbf{\Gamma} + \alpha_{dz} + \delta_{zt} + \gamma_s + \epsilon_{idzts} \quad (1)$$

where y_{idzts} is an outcome of individual i in district d, zone z, cohort t and survey year s. V_{dz} is our district-level measure of historical villagization as described in Section 3, and $treatedcohort_t$ is a dummy that equals 1 if individual i was born between 1960–1971 (0 if born between 1948–1959). α_{dz} denotes district fixed effects (which also include zone fixed effects), δ_{zt} denotes zone-cohort fixed effects, γ_s denotes survey year fixed effects and $\mathbf{X'}_{dz}$ is a vector of district-level controls. Our preferred specification includes the interaction between the pre-*Ujamaa* district-level primary schooling rate and the cohort dummy as a control variable. Other variables are controlled for in robustness checks, as discussed in Results Section 5. Baseline specification regressions are weighted by the survey weights provided in the data. We cluster standard errors at the district level.

Our coefficient of interest (β) is on the first interaction term and captures the average difference in outcomes between high- and low-villagization districts for individuals of political education age during the *Ujamaa* period. Alternatively, β represents the average difference in outcomes between individuals of political education age and other cohorts within a high-villagization district.[26] Since we do not directly observe individual exposure to villagization but villagization intensity at the district level, we interpret β as an intent-to-treat (ITT) effect, which is likely smaller than a treatment effect on the treated.

The baseline specification is appealing due to its simplicity and because pooling individuals into two cohorts increases statistical power. However, we also estimate a specification that allows treatment effects by cohorts to vary more flexibly over time. The main advantage of this second specification is that it allows us to assess pre-trends. In addition, we can test whether treatment effects are stronger for those that were of primary schooling age during

[26] In the baseline, we use a continuous measure of villagization rather than a high-/low-villagization dummy.

the entire villagization period than for those with only partial temporal overlap in exposure to the *Ujamaa* policy. The flexible specification is as follows:

$$y_{idzts} = \sum_{t=2}^{10} \beta_t (V_{dz} \cdot cohort_t) + \sum_{t=2}^{10} (\mathbf{X'}_{dz} \cdot cohort_t)\mathbf{\Gamma_t} + \alpha_{dz} + \delta_{zt} + \gamma_s + \epsilon_{idzts} \quad (2)$$

where $cohort_t$ is a dummy that indicates whether individual i belongs to cohort t. We divide our sample into ten 4-year cohorts (born between 1948-1951, 1952-1955, ..., 1984-1987). The cohort born between 1948 and 1951 ($t = 1$) is the omitted category. We interpret each of the parameters β_t as the impact of the *Ujamaa* policy on cohort t. Since villagization lasted from 1970 to 1981, we expect the coefficients for the cohorts born between 1960 and 1971 to be greater than 0 and the coefficients for the cohorts born before 1960 to be equal to 0. The coefficients for cohorts born after 1971 are somewhat ambiguous ex-ante but we expect them to decrease over time.[27]

4.2. Assumptions and Threats to Identification

The difference-in-differences estimate β in equation (1) can be interpreted as a causal ITT effect of the *Ujamaa* policy under the parallel trends assumption that, in the absence of the policy, the changes in the outcome variables across cohorts would not have been systematically different in low and high villagization districts within a zone. We now discuss the most important potential challenges to this and other assumptions.[28]

Level differences across districts. Districts with lower initial levels of national identity may have implemented the villagization policy more or less intensively. There may also be other district-level correlates with villagization and contemporary national identity, as discussed in Appendix A. Such level differences, insofar as they are constant across cohorts, are subsumed by district fixed effects and are thus not threats to identification given our approach.

National or regional policies and other differences between cohorts. The timing of the *Ujamaa* policy may have been correlated with other factors that shaped national identity across the country, such as macroeconomic developments, trends in education, or regional policies. Such factors, insofar as they affected the entire country or all districts

[27] The emphasis on nation building was removed from the official curriculum in 1992 (and likely received considerably less emphasis after villagization was officially abandoned in 1982). However, those who remained in the registered villages still likely enjoyed better access to public schooling than those in other areas, and the textbooks used in schools likely remained the same for a few years given a lack of alternatives.

[28] Moreover, we assume that the policy was unanticipated. We discuss potential violations of SUTVA (spillovers) in the robustness section. We are not concerned about the identification challenges related to difference-in-differences settings with staggered treatment highlighted in the recent econometrics literature (see Roth et al., 2023). The treatment in our main specification is not staggered and we do not estimate dynamic treatment effects.

within a subnational geographic zone similarly, are subsumed by cohort-zone fixed effects. Likewise, general differences in national identity between cohorts (for example due to age effects), insofar as they are constant within zones, are not a concern given the zone-cohort fixed effects.

Different trends over cohorts across districts. The development of national identity may have followed different pre-existing trends over cohorts across districts with differential intensity of villagization. For example, school cohorts in more remote districts, that were treated less (or more) intensively due to their distance from government institutions, may have had lower levels of national identity initially, and would have caught up with other districts even in the absence of the villagization policy. If not accounted for, such trend differences could violate the parallel trends assumption and thus pose a threat to identification (as opposed to level differences). The descriptive evidence in Figure 1 shows that this is unlikely to be a major concern, as low- and high-intensity districts follow parallel trends over cohorts that were of primary schooling age pre- and post-policy. We also look for and do not find signs of differential pre-trends more systematically in Section 5.

In addition, we interact cohort fixed effects with numerous pre-*Ujamaa* district characteristics that may be correlated with villagization intensity and that may affect national identity differently for different cohorts in the long run (see Appendix A). Our baseline includes the primary school enrollment rate in 1967 interacted with cohort fixed effects, which is significantly correlated with villagization intensity. In robustness checks, we also include several other controls in 1967 interacted with cohort fixed effects (see Section 5.4).

Migration. We observe respondents' districts at the time the contemporary surveys were conducted (2005–2008 for the main outcome variable) but not at the time of villagization (1970–1981). If a respondent lived in a different district when they were of primary schooling age vs. when they responded to a survey, this could bias the estimate of our coefficient of interest in different directions, depending on what determines migration. However, not all forms of migration threaten the validity of our results. First, note that within-district migration does not affect our results, as our measure of villagization varies at the district level. Second, if migration across districts is uncorrelated with villagization or national identity, this is akin to classical measurement error in the villagization variable, biasing our estimate towards zero. In that case, our coefficient estimate would be a lower bound on the true effect of the *Ujamaa* policy on outcomes. We are more concerned about selective migration possibly driving our results. In Section 5.4, we provide several pieces of evidence to show that selective migration is unlikely to explain our results.

5. Results: Ujamaa and National Identity

5.1. Descriptive Evidence

Before we discuss our main result, we show descriptive evidence on the levels of national identity by age cohort and intensity of villagization. Figure 1 plots the mean of our measure of national identity in 2005/2008 by birth cohort, relative to the level of the oldest cohort in our sample (born 1948-1951), for high- and low-villagization districts separately (above or equal to and below median villagization in the sample). For ease of interpretation, we standardize the measure of national identity to have a mean of 0 and a standard deviation of 1. We group respondents in 4-year cohorts to reduce noise.

We see few systematic differences across cohorts for low-villagization districts, where the *Ujamaa* policy was implemented to a lesser degree. In contrast, in high-villagization districts, national identity is higher on average for cohorts that were of schooling age during villagization compared to younger and older cohorts. Mean national identity of cohorts that were too old or too young during villagization (born 1948–1959 or 1972–1987) moves in parallel for high- and low-villagization districts, but diverges for cohorts that were of schooling age during villagization (those born 1960–1971). While this evidence is reassuring, we do not interpret it as causal as it does not account for the set of fixed effects and controls that we include in our main difference-in-differences specification, which we discuss next.

5.2. National Identity: Difference-in-differences Estimates

Table 2 shows our main difference-in-differences estimates of the ITT effect of the *Ujamaa* policy in 1970–1981 on national identity in 2005-2008 (β in equation (1)). Column (1) shows that national identity is on average 0.165 standard deviations higher for the treated cohort compared to the control cohort for every one standard deviation increase in villagization, controlling for district and cohort fixed effects, pre-*Ujamaa* primary schooling rates interacted with cohort fixed effects, and survey year fixed effects. Column (2), our preferred specification, additionally controls for zone-cohort fixed effects. Our main result in column (2) is that a one standard deviation increase in exposure to the *Ujamaa* policy increases national identity around two decades later by 0.226 standard deviations. The effect is statistically significant (p-value < 0.01). The difference between average expressed national identity of the treated and the control cohort, for a district with complete villagization compared to a district without any villagization, corresponds to the difference between a respondent "feeling more Tanzanian" and "feeling more a member of their ethnic group [than Tanzanian]".[29]

[29] As an additional exercise, in Appendix D and Table A.4, we explore and discuss the effect of *Ujamaa* on inter-ethnic marriage. We find that treated cohorts were more likely to marry across ethnic lines.

In columns (3) to (8), we additionally control for the pre-policy district characteristics indicated in the column heads interacted with the cohort dummy. These district characteristics are potential correlates with villagization discussed in Appendix A (distance to capital, district revenue, ethnolinguistic fractionalization, geographical features, hospital beds, and weather shocks). All coefficients on the interaction between villagization and the cohort dummy remain qualitatively similar and are significant at the 5% or 1% level. In the Robustness Section 5.4 further below, we include further controls interacted with the cohort dummy.[30]

5.3. Parallel Trends

Figure 2 displays the regression coefficients of our flexible specification including all cohorts (equation (2)). Each coefficient shows the differential effect of a one standard deviation increase in villagization on national identity for the birth cohort indicated on the x-axis compared to the reference cohort (born in 1948-1951). In line with our main result, we find positive effects of villagization on national identity for the cohorts that were of primary-school age during the policy (largest and significant at the 5% level for the cohort born in 1968–1971, i.e., in 4th grade age at the height of villagization) but not for older cohorts. We find a positive, albeit statistically insignificant effect on the cohort that entered primary-school age right after villagization officially ended. This is not surprising given that the schools established in the registered villages remained operational after the policy ended and political education was not abolished until 1992, although the zeal of its implementation likely waned after the end of villagization. For cohorts of primary school age longer after villagization ended, there is only a very small differential effect of villagization on national identity.

Figure 2 provides support for the parallel trends assumption. We see no differential pre-trends in national identity between high- and low-villagization districts over cohorts that were of primary schooling age before the policy. Nor do we see differential trends for cohorts of primary schooling age many years after the policy ended. If our main results were explained by unobserved factors correlated with villagization, they would need to have different effects for different cohorts following this specific pattern.

Appendix Table A.5 column (2) reports the point estimates and standard errors displayed in Figure 2. In addition, we re-estimate this specification, now controlling for the same variables as in our main Table 2, all interacted with cohort fixed effects. The coefficients of interest are qualitatively similar across all columns. The interaction for the birth cohort of 1968–1971, which was of schooling age at the height of villagization, is statistically significant at the 1% or 5% level in all columns except columns (1) and (8).

[30] Note that we cannot control for all district characteristics simultaneously, each interacted with cohort fixed effects, due to a lack of statistical power. However, in Table A.6 columns (8) and (9), we show robustness to controlling for the first principal component of all controls interacted with cohort fixed effects.

5.4. Robustness

Here we address several other potential empirical concerns. We return to our baseline specification in equation (1), which pools the treated cohorts born between 1960–1971 and compares them to the pooled cohorts born between 1948–1959. Tables 3 and A.6 show the results of these robustness checks for our preferred specification.

Additional controls. In Table 3 columns (3) to (10), we show that the main result presented in Table 2 is robust to controlling for further pre-*Ujamaa* district characteristics interacted with the cohort dummy. Columns (1) and (2) are identical to Table 2 for comparison. The additional controls include pre-*Ujamaa* levels of national identity based on a survey conducted in 1967 (Prewitt et al., 1970), urbanization, population density, GDP, the value of agricultural production (all from Jensen (1968)), distance to the border with Uganda, the presence of colonial missions (Hedde-von Westernhagen and Becker, 2022), and the total length of roads in the district in 1967 (Jedwab and Storeygard, 2022), respectively scaled by district population and area. The coefficients in columns (3) to (10) are qualitatively similar to those in the baseline in column (2) and highly statistically significant.[31] In Table A.6 columns (8) and (9), we show robustness to controlling for the first principal component of all controls interacted with cohort fixed effects.[32] These checks address concerns that the national identity of cohorts in districts differing in these characteristics followed differential trends for reasons unrelated to *Ujamaa*.

Migration. Migration during the time between villagization and the survey could bias our estimates since we only observe the current district of respondents at the time of survey and not where they lived during villagization. However, as discussed in Section 4.2, only selective migration threatens interpretation in a way that goes against our results. The main concern is that those with lower national identity were more likely to move to districts with lower villagization to evade the *Ujamaa* policy. Such migration could bias our estimates upwards because those with lower national identity would enter the control group. In Appendix B, we conduct two exercises to assess robustness of our findings to potential selective migration. Table A.6 column (5) drops all districts with either an in- or an out-migration rate in the highest deciles, with similar results to our baseline. As we explain in Appendix B, Columns (6) and (7) show that even under strong assumptions, selective migration could explain at most 36% of the effect of the *Ujamaa* policy on national identity we find. Hence, selective migration does not plausibly explain the majority of our coefficient given the migration rates in the data.

[31] The coefficient in column (3) is slightly smaller than our main estimate, which can be explained by the fact that the sample is different because data on national identity in 1967 is not available for some regions.

[32] Data on national identity in 1967 is not available in all districts. In column (8), we exclude this variable. In column (9), we include this variable with the caveat that the sample is different from the baseline.

Other robustness checks. In Appendix C, we show robustness to alternative regression weights, samples and functional forms, and address concerns about social desirability bias.

5.5. Channels

We conduct several exercises to examine two sets of channels that plausibly explain our results. The first, and our preferred explanation, is that *Ujamaa* primarily shaped the national identity of exposed cohorts through public education provided in planned villages. Alternatively, the outcomes we observe may reflect other cohort-specific factors that interacted with villagization. Note that the district fixed effects hold constant the average effect of villagization alone across all cohorts, so any channel must be specific to the cohort that was of the age to receive political education during *Ujamaa*. While villagization alone may have affected national identity across all cohorts, only those of political education age during *Ujamaa* were exposed to the bundled treatment (combining villagization with education). Hence, our main coefficient can be interpreted as the differential effect of the bundled treatment compared to the effect of villagization alone.[33]

To differentiate between the two sets of potential channels, we examine heterogeneous treatment effects by whether respondents completed primary school. In Table 4, we interact our treatment with a dummy that indicates whether the respondent completed formal primary school. We also control for the un-interacted schooling dummy as well as its interactions with the cohort dummy and the villagization variable. Column (1) indicates that the coefficient on the treatment is 0.331 standard deviations larger for those who completed primary school compared to those who did not. There is no significant effect of the treatment on national identity for individuals who did not complete formal primary school. This result is consistent with public schooling as the primary channel through which the *Ujamaa* policy shaped national identity, rather than other differences between age cohorts that may have interacted with villagization.[34]

That the effect of *Ujamaa* on national identity is stronger for males, who had higher rates of school attendance than females (Chamie, 1983) (Table 4 columns (3) and (4)), corroborates the importance of schooling.

Having established public education as the most plausible channel through which the *Ujamaa* policy influenced national identity, we turn to an exploration of more specific

[33] The effect of villagization alone appears to be small. Levels of national identity are no higher in more intensively villagized districts for cohorts outside of schooling age during the *Ujamaa* era (see Figure 1).

[34] An alternative explanation for this result is selection into primary school: it is theoretically possible that parents opposed to the regime's effort to indoctrinate their children were less likely to send their children to school during *Ujamaa*. If these children report less strong feelings of national identity in later surveys, this selection rather than the effect of attending primary school could explain the heterogeneity result. However, such selection into schooling is unlikely to be an important channel underlying our main result on national identity given that the majority of children in registered villages were enrolled in public school.

mechanisms. First and foremost, the content of the new political education curriculum was clearly oriented toward establishing national identity, as we outline in Section 2 (Background). As detailed below, we find empirical support for this mechanism as the primary driver of the public education channel. We also explore the role of establishing Swahili as the language of instruction. Additionally, we consider that public education may have facilitated the formation of a national identity through inter-group contact across ethnic lines, and potential implications for human capital, income, and occupational choice. These and other mechanisms are examined in the sections below.

Public education curriculum. Villagization was intentionally disorienting, and served to disrupt traditional norms and networks. Subsequent reorientation – away from distinct, ethnic identities, and toward a consolidated, national identity – was facilitated primarily through the primary school curriculum. Education reform was arguably more effective than concurrent policies as it targeted citizens during particularly "impressionable years" (Krosnick and Alwin, 1989).

Changes to the curriculum were implemented alongside a nation-wide increase in the supply of education. However, our analysis suggests that changes in the *content* of education that was delivered under the *Ujamaa* policy in the registered villages, rather than a general increase in the *supply* of education, explain our results.[35] Recall from Figure 2 that we see few differential effects of villagization on cohorts who were of primary schooling age after the end of *Ujamaa* – and especially for cohorts that entered primary schooling age after the abolition of Political Education in 1992. Moreover, the effect of villagization on school completion for the treated cohort specifically was modest in size (Table A.7, column (1)).[36]

We further isolate the impact of the revised curriculum by ruling out the possibility that the policy served to improve education *quality*. First, if the quality of education were higher among treated cohorts, we would expect to see increased human capital in those cohorts. However, as we discuss below, we find no positive or significant effects on relevant outcomes such as occupational choice and income. Second, the directive to overhaul the curriculum was not necessarily accompanied by additional resources.[37] Third, one of Nyerere's goals with the Education for Self-Reliance policy was to "reduce elitism and the tendency for schooling

[35] The overall supply of public education increased for all cohorts, even after the *Ujamaa* policy ended, since the public schooling infrastructure remained. However, the elements of the policy that explicitly targeted nation building waned with the policy's repeal.

[36] On the one hand, enrollment increased nationwide; on the other hand, as documented in Appendix A, pre-*Ujamaa* enrollment, which we control for, is positively correlated with villagization. This is likely due to the villagization campaign being easier to implement in places that had existing infrastructure, which in turn weakens the differential effect of villagization on primary school completion.

[37] While Nyerere's post-independence reforms included additional investments (a general expansion of public goods and services including water, land, agricultural and veterinary production supplies and equipment), which served in part as inducements to encourage resettlement into the *Ujamaa* villages (McHenry, 1979), these seem unlikely to have had cohort-specific effects on national identity.

to further social and other inequalities and class formation." Thus it seems unlikely that the results we observe are driven by an increase in school quality, but rather, the content of the new curriculum.

Swahili as the national language of instruction. In addition to revising the content of the public school curriculum, the *Ujamaa* era saw the establishment of Swahili as the national language of instruction in all primary schools. Around the world, the establishment of a common, national language has played an important role in strengthening national identity (see e.g., Alesina et al., 2021). To see whether this is the case in our setting, we examine whether treated individuals are more likely to be proficient in Swahili. Appendix Table A.7 column (2) shows the effect of our treatment on literacy in Swahili based on census 1988 data from IPUMS. The effect is positive and statistically significant but small. This is likely due to the fact that the overall level of Swahili proficiency is very high – both in Tanzania and in our sample. According to Afrobarometer Round 4, only 8 out of 1,208 respondents in our sample report not speaking Swahili well, and 99.0% of respondents in Round 3 and 93.7% of respondents in Round 4 report their home language being from the Bantu language family (a group of languages that are linguistically very similar; Swahili is the most common Bantu language). Moreover, all Afrobarometer interviews were conducted in Swahili, so all respondents must be proficient.

Taken together, while Swahili as a national language may have helped with nation building in Tanzania, everyone's similar level of Swahili proficiency means it's unlikely to explain why *Ujamaa* had a different impact on national identity among certain groups.

Intergroup contact. An alternative and perhaps complementary channel through which the *Ujamaa* policy might have shaped national identity is through intergroup contact across ethnic lines. Intergroup contact has a long history in the social sciences, beginning with psychologist Gordon Allport (1954) specifying the conditions under which it can reduce prejudice. Allport's (1954) hypothesis has been confirmed in more recent meta-studies (Hewstone et al., 2014) and economists have also begun to more rigorously identify its effects – for example showing that collaborative intergroup contact between individuals from different social groups may foster cooperation and reduce the salience of group identity (Bazzi et al., 2019; Rao, 2019; Lowe, 2021). In the context of villagization, bringing children from different ethnic groups together in public schools in *Ujamaa* villages could have decreased the salience of ethnic identity in favor of national identity. However, scholars have also shown that intergroup contact can also foster exclusionary attitudes (Enos, 2014) and challenge social solidarity (Putnam, 2007). In our context, it is also possible that intergroup contact *sharpened* the salience of ethnic identities, for example due to intergroup competition for limited resources in the villages.

If intergroup contact played an important role in explaining the effect of our treatment on national identity, we would expect larger effects in places where such contact was more

likely to take place. To test this idea, we interact our treatment (cohort interacted with villagization) with district-level ethnolinguistic fractionalization (ELF) in 1967.[38] We also control for the lower-level interactions that are not included in the fixed effects. Table 4 column (2) shows the results for our measure of national identity. The triple interaction of villagization × cohort × ELF is small and statistically insignificant.

In addition, as discussed in Section 6 below, we note that *Ujamaa* did not lead to an increase in generalized trust, as would be consistent with intergroup contact as the main channel underlying our results.

Occupational choice. Schooling under *Ujamaa* not only included political education but may have also provided students with enhanced human capital and skills. Resulting occupational choices may have in turn bolstered national identity. For example, those who obtained public education during the *Ujamaa* period may be more likely to work in public sector jobs. We investigate this in Appendix Table A.8 which is analogous to our baseline specification but with dummies for different occupations as outcome variables (employed in the government including parastatal, employed in the private sector, agriculture or self-employed in other sectors).[39] All coefficients are neither sizeable nor statistically significant, indicating that the treatment had little to no effect on occupational choice.

Economic effects. As explained above, it is possible that education under *Ujamaa* led to improved human capital or income. If schooling increased individuals' earnings, which in turn fostered a sense of gratitude towards the new Tanzanian nation-state, this may be an alternative explanation for our finding on the effect of *Ujamaa* on national identity.

At the outset, it is important to note that *Ujamaa* likely had negative economic effects on individuals on average (Collier, 1988). If anything this goes against our main finding on national identity if economic well-being and national identity are positively correlated. Economic effects are thus unlikely to play a major role in driving our findings. Nevertheless, we proceed to test this potential mechanism by examining the economic effects of *Ujamaa* using various measures of income. Appendix Table A.9 columns (1) to (4) show the results for household consumption and expenditures from the Tanzania National Panel Survey 2008 (as above), as well as for a dummy that equals 1 if an individual has a job that earns a cash income and a wealth index constructed from Afrobarometer data (Afrobarometer does not contain consumption data). We find that treated individuals have lower measures of income and wealth in 2005/2008 compared to the control cohort. That is, schooling under *Ujamaa* appears to have coincided with an overall *decrease* in income. In addition, columns (5) and

[38] ELF is computed as 1 minus the Herfindahl concentration index of ethnolinguistic group shares in each district given by the 1967 population census data. A higher ELF index corresponds to a higher degree of ethnic diversity within a district.

[39] The data is from the Tanzania National Panel Survey Round 1 (2008). We use this data, which also contains the birth years and district of respondents, since information on occupation is available for only a small subsample of respondents in the Afrobarometer data.

(6) show that national identity and the proxies of economic well-being from Afrobarometer are weakly positively correlated. These results together mean that *Ujamaa* if anything had negative economic effects that translated into weakly negative effects on national identity. Hence, this potential channel does not plausibly explain our results on national identity.

The end of *Ujamaa*. A final question is whether the external developments that led to the end of *Ujamaa*, including a global economic crisis and the invasion by Uganda, had cohort-specific effects that explain our results. The short answer is no. We elaborate on our answer and provide empirical tests in Appendix E.

6. Results: *Ujamaa* and State Legitimacy

We now address the question of whether, in building the Tanzanian *nation*, *Ujamaa* strengthened the nascent Tanzanian *state*. This is an important question given the ambitions of Nyerere and other state builders throughout history to use nation building as a way to overcome societal divisions and establish the state as a legitimate, central authority. Our results point to success in this regard, but also highlight important trade-offs. As we detail below, treated cohorts are more likely to express attitudes in favor of a strong, central state. However, the results also suggest that efforts to build a strong state through *Ujamaa* engendered acquiescence to authoritarianism.

In what follows, we examine the impact of *Ujamaa* on a range of attitudes toward the state, including respect for authority, support for one-party rule, and trust in government and government-run media. We compare the latter to trust in independent media and generalized trust. We also look at attitudes and actions related to citizens' engagement with the state and state institutions. All specifications follow the same empirical strategy outlined in Section 4, replacing national identity with different outcome variables from Afrobarometer. Figure 3 illustrates the results. We show the corresponding coefficients in Appendix Table A.10. As above, we report standardized coefficients to facilitate interpretation for all non-binary outcomes. The coefficient of interest is the interaction between the respondent's district-level measure of villagization and a dummy that indicates whether the respondent is in the treated cohort. We interpret this coefficient as the effect of the *Ujamaa* policy on the outcomes stated on the y-axis in the Figure (column heads in the table).

The first outcome is based on a question asking respondents which one of two statements regarding views of the state is closest to their views: *Statement 1: Citizens should be more active in questioning the actions of leaders* or *Statement 2: In our country, citizens should show more respect for authority.* We find that exposure to *Ujamaa* has a sizable and statistically significant effect on respect for state authority: a one standard deviation increase in the treatment increases the outcome variable by 0.169 standard deviations ($p < 0.05$). Prior studies have interpreted this variable as measuring citizens' critical attitudes

toward government (Doorenspleet, 2012), which has also been understood as an indicator of demand for democracy (Inglehart, 1997). We interpret these results as suggesting that citizens who were exposed to the *Ujamaa* policy are more likely to see the state as a legitimate central authority, but less likely to express demand for democracy.

The second outcome examines respondents' stated approval for a system of government where only one political party is allowed to stand for election and hold office. We find that exposure to the *Ujamaa* policy is positively correlated with approval of one-party rule: a one standard deviation increase in the treatment is associated with a 0.097 standard deviations higher measure of approval of one-party rule, although the coefficient is imprecisely estimated ($p = 0.138$). This is perhaps unsurprising, given that Nyerere saw one-party rule as necessary to foster national integration in a country characterized by substantial ethnic differences (Komba, 1996). Our results speak to the legacy of one-party rule, as well as the intertwined nature of the party and the state, which has persisted despite the (re-) introduction of multi-partyism in 1992 (Paget, 2021).

The third outcome is based on a question asking respondents which of two statements is closest to their views: *Statement 1: People are like children; the government should take care of them like a parent* or *Statement 2: Government is like an employee; the people should be the bosses who control the government.* We find that exposure to *Ujamaa* has a sizable and statistically significant effect on paternalistic attitudes: a one standard deviation increase in the treatment is estimated to increase the outcome variable by 0.230 standard deviations ($p < 0.05$). Such acquiescence to government authority can facilitate voluntary compliance with state policies. Scholars point to extensive buy-in of public health measures such as malaria control (Croke, 2012) among Tanzanians – in contrast to citizens of neighboring states. Such compliance is difficult to achieve without acceptance of the state as a legitimate authority.

For the fourth and fifth outcomes, we calculate the difference between stated trust in government newspapers (TV/radio) and in independent newspapers (TV/radio). Exposure to the *Ujamaa* policy is positively correlated with stated trust in government broadcasting compared to independent media (0.097 and 0.205 standard deviations).

The sixth outcome is a measure of citizen engagement, capturing whether respondents report getting together with others to raise an issue. We interpret this as another proxy for critical attitudes toward government. The negative and only marginally insignificant coefficient (-5.1 percentage points ($p = 0.106$)) is in line with the result for the first outcome (respect for state authority), in that it indicates greater acceptance of the state as a legitimate authority.

Appendix Table A.11 presents a number of additional results to examine how *Ujamaa* shifted attitudes related to state legitimacy. The first provides evidence consistent with *Ujamaa* strengthening citizens' expectations and preferences for public goods provision by the state: column (1) shows the coefficient of a regression of support for free schooling on our

treatment (analogous to our main specification). The coefficient is positive and statistically significant at all conventional levels. The second set of results considers a set of placebo outcomes: it shows that *Ujamaa* appears to have little effect on gender norms (columns (2) and (3)). In sum, these results show that *Ujamaa* had a persistent effect on citizens' preferences along the dimensions it explicitly sought to influence.

Finally, we examine whether exposure to *Ujamaa* influences contemporary political support for the ruling party, which has been in power in different forms since independence and is thus the heir to Nyerere's TANU regime. First, Appendix Table A.11 column (4), which is based on Afrobarometer data and exploits variation across districts and cohorts, shows that *Ujamaa* had a small positive effect on respondents reporting feeling close to the ruling CCM party.[40] Second, we analyze voting patterns at the district level using electoral data. Note that we cannot conduct the analysis using our standard specification because electoral data is only available at the constituency level but not by age cohorts. As we show in column (5), there is a positive and statistically significant correlation between historical villagization in a district and support for the CCM in 2000-2005 (proxied by turnout multiplied with the CCM presidential candidate's vote share in the national elections),[41] primary school enrollment rate in 1967, and zone fixed effects (analogously to our main empirical strategy).[42] The results are consistent with *Ujamaa* fostering support for the ruling CCM party in the long run, and also dovetail with the results showing support for one-party rule among the treated cohort.

In sum, we find that cohorts exposed to *Ujamaa* are more likely to respect state authority and approve of one-party rule, and have higher trust in central government institutions such as state media. These findings imply that the *Ujamaa* policy contributed to establishing the new Tanzanian state as a legitimate central authority.

There is evidence to suggest that a strong shared identity and trust in a commonly accepted central state are important ingredients for avoiding fragility and civil conflict (Besley, 2020). Within Africa, intra-state conflict has been more prevalent in countries where a smaller proportion of the population identifies with the nation as a whole (Besley and Reynal-Querol, 2014), as we show in Appendix Figure A.6. This correlation is consistent with our narrative that creating a strong national identity may contribute to loyalty to the state and more political stability as a result. As we show in Appendix Figure A.7, countries with high ethnic diversity (as measured by ethnolinguistic fractionalization) tend to have a higher incidence of internal conflict. In contrast, Tanzania today is a clear outlier, with a very low prevalence

[40] The estimates are statistically insignificant at conventional levels. This could be partly explained by there being little multi-party competition in the post-*Ujamaa* years or by reporting bias.

[41] Turnout is a commonly used proxy for support for the ruling party in a hegemonic party system.

[42] Data on 2000-2005 voting are from Carlitz (2017) and are based on data from the National Electoral Commission.

of internal conflict despite being one of the most ethnically diverse countries.[43]

However, we note that efforts to forge a strong, shared national identity do not necessarily generate more cooperation and generalized trust among the population. While exposure to the *Ujamaa* policy generated higher levels of trust in state institutions, it did not have the same impact on generalized intra- or inter-ethnic trust. As we show in Figure 3 and Appendix Table A.10, we find only small and statistically insignificant effects of the *Ujamaa* policy on trust in members of one's own ethnic group (0.055 standard deviations) or in members of other ethnic groups (-0.141 standard deviations). If anything, the *Ujamaa* policy generally *decreased* inter-ethnic trust.[44]

In sum, the results we obtain for attitudes toward the state and generalized inter-ethnic trust indicate that the creation of a national identity in a top-down manner first and foremost strengthened the one-party state's ability to govern but that the extent of social cohesion across ethnic groups remained limited.

7. Conclusion

Leaders throughout history have attempted to overcome the challenges of governing diverse populations by using 'bundles' of nation-building measures, including public education and resettlement. We study the consequences of one of the largest such efforts in post-colonial Africa — the Tanzanian *Ujamaa* policy — which combined the resettlement of millions of people with a public education reform. We find individuals most affected by the policy are more likely to primarily identify as Tanzanian rather than with their ethnic group. The effects are persistent and substantive.

Ujamaa's positive impact on national identity stands in contrast to the results of previous studies of state-building policies involving forced resettlement of diverse populations, which document that such efforts frequently resulted in inter-group conflict (e.g., Dippel, 2014). The contrasting result for *Ujamaa* arguably reflects the policy's bundled nature: by combining resettlement with an education reform, the nascent Tanzanian state was able to project its ideology to a similarly young and impressionable population.

We also find that *Ujamaa* helped the new Tanzanian state establish itself as a legitimate central authority. At the same time, the policy reduced demand for government account-

[43] We leave it to future research to further test this narrative by examining the direct impact of *Ujamaa* on conflict prevalence within Tanzania. Data on conflict would need to be at the individual or cohort level given our identification strategy.

[44] This result may at first seem at odds with our findings for ethnic intermarriage in the Appendix. However, we should not necessarily understand intermarriage as an expression of inter-ethnic, or generalized trust. Rather, intermarriage reflects beliefs one has about a particular individual they marry from a certain group, versus beliefs about a random member of that group. Thus, we should understand intermarriage as an expression of *personalized* trust facilitated by repeated bilateral interactions. Generalized and personalized trust need not be correlated (Guiso et al., 2009).

ability and engendered greater preferences for non-democratic governance, highlighting the difficult balance inherent to effective state building (Acemoglu and Robinson, 2020). Members of the treated cohort are more likely to trust state institutions and less likely to question state authority. However, we find few signs of increased cooperation and generalized trust in other ethnic groups among those most exposed to *Ujamaa*.

Our examination of *Ujamaa* has broader implications for the study of state building in diverse societies, and the sub-field of education and governance in particular. Since Dewey (1916), prominent scholars have argued that education cultivates a "culture of democracy," driving political legitimacy and societal trust (Lipset, 1959; Putnam et al., 1994). In contrast, theories of nation building (Hobsbawm, 1992; Tilly and Ardant, 1975) imply that mass education is primarily used as a vehicle of indoctrination – by autocracies as well as democracies. Our findings suggest a need for nuance: we find that while state education can successfully foster national identity, such efforts may not fundamentally shift attitudes towards higher societal trust and preference for democracy.

Our finding that efforts to build a national identity can promote relatively unquestioning support for state authority raises important questions related to the potentially anti-democratic nature of state-building reforms. These questions are particularly important in light of widespread concern about democratic backsliding and institutional erosion around the world (Hyde, 2020). Our results encourage further scrutiny of the political economy of state building in diverse societies. We demonstrate that the choices leaders make on how to build a shared identity from diverse groups are first order and deserve further attention.

Data Availability Statement. The data and code underlying this article is available on Zenodo at https://dx.doi.org/10.5281/zenodo.13896187

References

Abramenko, S., V. Korovkin, and A. Makarin (2024). Social Media as an Identity Barometer: Evidence from the Russia-Ukraine War. *AEA Papers and Proceedings*.

Acemoglu, D. and J. A. Robinson (2020). *The Narrow Corridor: States, Societies, and the Fate of Liberty*. Penguin Books.

Afrobarometer (2017). *Afrobarometer Data, Tanzania*. http://www.afrobarometer.org.

Alesina, A., R. Baqir, and W. Easterly (1999). Public goods and ethnic divisions. *The Quarterly Journal of Economics 114*(4), 1243–1284.

Alesina, A., W. Easterly, and J. Matuszeski (2011). Artificial states. *Journal of the European Economic Association 9*(2), 246–277.

Alesina, A., P. Giuliano, and B. Reich (2021). Nation-building and education. *The Economic Journal 131*(638), 2273–2303.

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

Alesina, A. and E. Spolaore (1997). On the number and size of nations. *The Quarterly Journal of Economics 112*(4), 1027–1056.

Allport, G. W. (1954). *The Nature of Prejudice*. Reading, MA: Addison-Wesley.

Anderson, B. (1991). *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Verso.

Assouad, L. (2021). Charismatic leaders and nation-building. *Working Paper, Paris School of Economics*.

Bahry, D., M. Kosolapov, P. Kozyreva, and R. K. Wilson (2005). Ethnicity and trust: Evidence from russia. *American Political Science Review 99*(4), 521–532.

Bandiera, O., M. Mohnen, I. Rasul, and M. Viarengo (2019). Nation-building through compulsory schooling during the age of mass migration. *The Economic Journal 129*(617), 62–109.

Bazzi, S., A. Gaduh, A. D. Rothenberg, and M. Wong (2019). Unity in diversity? How intergroup contact can foster nation building. *American Economic Review 109*(11), 3978–4025.

Bazzi, S., M. Hilmy, and B. Marx (2023). Religion, education, and the state. *NBER Working Paper*.

BenYishay, A., R. Rotberg, J. Wells, Z. Lv, S. Goodman, L. Kovacevic, and D. Runfola (2017). *Geocoding Afrobarometer Rounds 1 - 6: Methodology & Data Quality*. Williamsburg, VA: AidData.

Besley, T. (2020). State capacity, reciprocity, and the social contract. *Econometrica 88*(4), 1307–1335.

Besley, T. and T. Persson (2011). *Pillars of Prosperity*. Princeton University Press.

Besley, T. and M. Reynal-Querol (2014). The legacy of historical conflict: Evidence from Africa. *American Political Science Review 108*(2), 319–336.

Blanc, G. and M. Kubo (2021). Schools, language, and nations: Evidence from a natural experiment in France. *Brown University*.

Blouin, A. and S. W. Mukand (2019). Erasing ethnicity? Propaganda, nation building, and identity in Rwanda. *Journal of Political Economy 127*(3), 1008–1062.

Bluhm, R., R. Hodler, and P. Schaudt (2021). Ethnofederalism and ethnic voting. *CESifo Working Paper*.

Brautigam, D. (2023, February). The developing world's coming debt crisis. *Foreign Affairs*.

Bryceson, D. F. (1982). Peasant commodity production in post-colonial Tanzania. *African Affairs 81*(325), 547–567.

Bureau of Statistics, Ministry of Planning and Economic Affairs (1981). *1978 Population Census*. Dar es Salaam: United Republic of Tanzania.

Bureau of Statistics/Tanzania and Macro International (1997). *Tanzania Demographic and Health Survey 1996*. Calverton, Maryland: Bureau of Statistics/Tanzania and Macro International.

Burgess, R., R. Jedwab, E. Miguel, A. Morjaria, and G. Padró i Miquel (2015). The value of democracy: Evidence from road building in Kenya. *American Economic Review 105*(6), 1817–51.

Cantoni, D., Y. Chen, D. Y. Yang, N. Yuchtman, and Y. J. Zhang (2017). Curriculum and ideology. *Journal of Political Economy 125*(2), 338–392.

Cantoni, D. and N. Yuchtman (2013). The political economy of educational content and development: Lessons from history. *Journal of Development Economics 104*, 233–244.

Carlitz, R. D. (2017). Money flows, water trickles: Understanding patterns of decentralized water provision in Tanzania. *World Development 93*, 16–30.

Central Statistical Bureau, Ministry of Economic Affairs and Development Planning (1968). *1967 Population Census*. Dar es Salaam: United Republic of Tanzania.

Chamie, M. (1983). *National, institutional, and household factors affecting young girls' school attendance in developing societies*. International Center for Research on Women.

Cinnirella, F. and R. Schueler (2018). Nation building: The role of central spending in education. *Explorations in Economic History 67*, 18–39.

Clingingsmith, D., A. I. Khwaja, and M. Kremer (2009). Estimating the impact of the Hajj: Religion and tolerance in Islam's global gathering. *The Quarterly Journal of Economics 124*(3), 1133–1170.

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

Collier, D., P. Collier, S. Radwān, S. Radwān, S. Wangwe, S. Radwan, A. Wagner, P. mondial de l'emploi, I. L. Office, and W. E. Programme (1986). *Labour and Poverty in Rural Tanzania: Ujamaa and Rural Development in the United Republic of Tanzania*. Clarendon Press.

Collier, P. (1988). Macro-economic policy, employment and living standards in Malawi and Tanzania, 1973-84. Technical report.

Costello, M. J. (1996). Administration triumphs over politics: The transformation of the Tanzanian state. *African Studies Review 39*(1), 123–148.

Croke, K. (2012). The political economy of child mortality decline in Tanzania and Uganda, 1995–2007. *Studies in Comparative International Development 47*(4), 441–463.

Dehdari, S. H. and K. Gehring (2022). The origins of common identity: Evidence from Alsace-Lorraine. *American Economic Journal: Applied Economics 14*(1), 261–92.

Dell, M. and P. Querubin (2018). Nation building through foreign intervention: Evidence from discontinuities in military strategies. *The Quarterly Journal of Economics 133*(2), 701–764.

Depetris-Chauvin, E., R. Durante, and F. Campante (2020). Building nations through shared experiences: Evidence from African football. *American Economic Review 110*(54), 1572–1602.

Dewey, J. (1916). Nationalizing education. *Journal of Education 84*(16), 425–428.

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

Doorenspleet, R. (2012). Critical citizens, democratic support and satisfaction in African democracies. *International Political Science Review 33*(3), 279–300.

Duflo, E. (2001). Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. *American Economic Review 91*(4), 795–813.

Duggan, C. (2008). *The Force of Destiny: A History of Italy Since 1796*. Houghton Mifflin Harcourt.

Easterly, W. and R. Levine (1997). Africa's growth tragedy: Policies and ethnic divisions. *The Quarterly Journal of Economics 112*(4), 1203–1250.

Enos, R. D. (2014). Causal effect of intergroup contact on exclusionary attitudes. *Proceedings of the National Academy of Sciences 111*(10), 3699–3704.

Ergas, Z. (1980). Why did the Ujamaa village policy fail? Towards a global analysis. *The Journal of Modern African Studies 18*(3), 387–410.

Erikson, E. H. (1968). *Identity, Youth and Crisis*. W.W. Norton & Company.

Feierman, S. (1990). *Peasant Intellectuals Anthropology and History in Tanzania*. University of Wisconsin Press, Madison.

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

Gehring, K. (2022). Can external threats foster a European Union identity? Evidence from Russia's invasion of Ukraine. *The Economic Journal 132*(644), 1489–1516.

Gellner, E. (2006). *Nations and Nationalism*. New perspectives on the past. Blackwell.

Giuliano, P., A. Riboni, and B. Reich (2023). What motivates leaders to invest in nation building? In *Nation Building: Big Lessons from Successes and Failures*. Paris & London: CEPR Press.

Government of Tanzania (1984). Universal primary education and village based teacher training programme in Tanzania. Ministry of Education.

Guiso, L., P. Sapienza, and L. Zingales (2009). Cultural biases in economic exchange? *The Quarterly Journal of Economics 124*(3), 1095–1131.

Hedde-von Westernhagen, C. and B. Becker (2022). Mapping Missions: New Data for the Study of African History: Social and Economic History. *Research Data Journal for the Humanities and Social Sciences 7*(1), 1–33.

Herbst, J. (2014). *States and Power in Africa: Comparative Lessons in Authority and Control*. Princeton University Press.

Hewstone, M., S. Lolliot, H. Swart, E. Myers, A. Voci, A. Al Ramiah, and E. Cairns (2014). Intergroup contact and intergroup conflict. *Peace and Conflict: Journal of Peace Psychology 20*(1), 39.

Hobsbawm, E. J. (1992). *Nations and Nationalism since 1780: Programme, Myth, Reality*

(2 ed.). Canto. Cambridge University Press.

Holsti, K. J. (1996). *The State, War, and the State of War*. Cambridge University Press.

Hyde, S. D. (2020). Democracy's backsliding in the international environment. *Science 369*(6508), 1192–1196.

Hydén, G. (1980). *Beyond Ujamaa in Tanzania: Underdevelopment and an Uncaptured Peasantry*. University of California Press.

Inglehart, R. (1997). *Modernization and Postmodernization: Cultural, Economic, and Political Change in 43 Societies*. Princeton University Press.

Jarotschkin, A. (2018). *Experiences historiques et développement économique*. Ph. D. thesis, PSL Research University.

Jedwab, R. and A. Storeygard (2022, 06). The Average and Heterogeneous Effects of Transportation Investments: Evidence from Sub-Saharan Africa 1960–2010. *Journal of the European Economic Association 20*(1), 1–38.

Jensen (1968). *Government of Tanzania District Statistics*. Dar es Salaam: Government Printers.

Kaiser, P. J. (1996). Structural adjustment and the fragile nation: the demise of social unity in Tanzania. *The Journal of Modern African Studies 34*(2), 227–237.

Kikula, I. S. (1997). *Policy Implications on Environment: The Case of Villagisation in Tanzania*. Nordic Africa Institute.

Komba, W. L. M. (1996). *Changing Politics and Political Culture in Tanzania: The Impact on Political Education and Civics Curricula 1967-1994*. Ph. D. thesis, Institute of Education, University of London.

Krosnick, J. A. and D. F. Alwin (1989). Aging and susceptibility to attitude change. *Journal of personality and social psychology 57*(3), 416.

Lipset, S. M. (1959). Some social requisites of democracy: Economic development and political legitimacy. *American Political Science Review 53*(1), 69–105.

Lorgen, C. C. (2000). Villagisation in Ethiopia, Mozambique, and Tanzania. *Social Dynamics 26*(2), 171–198.

Lowe, M. (2021). Types of contact: A field experiment on collaborative and adversarial caste integration. *American Economic Review 111*(6), 1807–44.

Mamdani, M. (2003). Making sense of political violence in postcolonial Africa. In *War and Peace in the 20th Century and Beyond*, pp. 71–99. World Scientific.

Martin, T. D. (2001). *The Affirmative Action Empire: Nations and Nationalism in the Soviet Union, 1923-1939*. Cornell University Press.

Mbunda, F. (1982). Primary education since 1961. In *The Tanzanian experience: Education for liberation and development*, Volume 16. Unesco Institute for Education.

McHenry, D. E. (1979). *Tanzania's Ujamaa Villages. The Implementation of a Rural Development Strategy*. Institute of International Studies, University of California, Berkeley.

Michalopoulos, S. and E. Papaioannou (2013). National Institutions and Subnational Development in Africa. *The Quarterly Journal of Economics 129*(1), 151–213.

Michalopoulos, S. and E. Papaioannou (2016). The long-run effects of the Scramble for Africa. *American Economic Review 106*(7), 1802–48.

Miguel, E. (2004). Tribe or nation? Nation building and public goods in Kenya versus Tanzania. *World Politics 56*(3), 327–362.

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

Minnesota Population Center (2015). Integrated public use microdata series, international: Version 6.4.

National Bureau of Statistics (2010). *Tanzania National Panel Survey - Wave 1, 2008*. Dar es Salaam, Tanzania: NBS.

Ngallaba, S., I. R. Saidi Hussein Kapiga, and J. T. Boerma (1993). *Tanzania Demographic and Health Survey 1991/1992*. Dar es Salaam, Tanzania: Bureau of Statistics/Tanzania and Macro International.

Nyerere, J. (1963). *Democracy and the Party System*. Tanganyika Standard.

Nyerere, J. K. (1969a). *Freedom and Socialism/Uhuru na Ujamaa: A Selection from Writings and Speeches, 1965-1967*. Oxford University Press.

Nyerere, J. K. (1969b). *Ujamaa - Essays on Socialism*. Oxford University Press.

Nyerere, J. K. (1982). *Five Years of CCM Government*. Dar es Salaam, Tanzania: CCM.

Okunogbe, O. (2024). Does exposure to other ethnic regions promote national integration? Evidence from Nigeria. *American Economic Journal: Applied Economics 16*(1), 157–192.

Omari, C. K. (1995). The management of tribal and religious diversity. In *Mwalimu: Influence of Nyerere*. London: James Currey.

Osafo-Kwaako, P. (2012). *Essays in Economic History and Development*. Ph. D. thesis, Harvard University.

Paget, D. (2021). Tanzania: the authoritarian landslide. *Journal of Democracy 32*(2), 61–76.

Peeples, A. (2018). Relearning Ujamaa: Education in Tanzania from 1954 to 2002. *Department of History, UNC Thesis*.

Prewitt, K., G. Von Der Muhll, and D. Court (1970). School Experiences and Political Socialization: A Study of Tanzanian Secondary School Students. *Comparative Political Studies 3*(2), 203–225.

Putnam, R. D. (2007). E Pluribus Unum: Diversity and Community in the Twenty-first Century The 2006 Johan Skytte Prize Lecture. *Scandinavian Political Studies 30*(2), 137–174.

Putnam, R. D., R. Leonardi, and R. Y. Nanetti (1994). *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton University Press.

Putterman, L. (1986). *Peasants, Collectives, and Choice: Economic Theory and Tanzania's Villages*. Greenwich CT: Jai Press Inc.

Rao, G. (2019). Familiarity does not breed contempt: Generosity, discrimination, and diversity in Delhi schools. *American Economic Review 109*(3), 774–809.

Ray, D. and J. Esteban (2017). Conflict and development. *Annual Review of Economics 9*,

263–293.

Roberts, G. (2014). The Uganda–Tanzania War, the fall of Idi Amin, and the failure of African diplomacy, 1978–1979. *Journal of Eastern African Studies 8*(4), 692–709.

Rohner, D. and E. Zhuravskaya (Eds.) (2023). *Nation building: Big lessons from successes and failures*. CEPR Press.

Roth, J., P. H. Sant'Anna, A. Bilinski, and J. Poe (2023). What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics*.

Scott, J. C. (1998). *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Yale University Press.

Sitari, T. (1983). Settlement changes in the Bagamoyo district of Tanzania as a consequence of villagization. *Fennia-International Journal of Geography 161*(1), 1–90.

Smith, A. D. (1986). State-making and nation-building. *States in History 15*, 228–263.

Sturmer, M. (1998). *The Media History of Tanzania*. Ndanda Mission Press.

Tilly, C. and G. Ardant (1975). *The Formation of National States in Western Europe*. Princeton University Press.

Tripp, A. M. (1999). The political mediation of ethnic and religious diversity in Tanzania. In *The Accommodation of Cultural Diversity*, pp. 37–71. Springer.

Weber, E. (1976). *Peasants into Frenchmen: the Modernization of Rural France, 1870-1914*. Stanford University Press.

Wimmer, A. (2018). *Nation Building: Why Some Countries Come Together While Others Fall Apart*. Princeton University Press.

Wu, C. and Z. Shi (2020). Education and social trust in transitional china. *Chinese Sociological Review 52*(2), 115–143.

TABLES AND FIGURES

Table 1: Descriptive Statistics

Variable	Data Source	Mean	Std.Dev.	Min.	Max.	Obs.
Panel A. District level, 1967 borders (baseline sample)						
Villagization (share of rural population in registered villages in 1978)	Population Census	0.95	0.08	0.52	1.00	52
Primary school enrollment rate in 1967 (per '000 inhabitants)	Jensen (1968)	70.90	23.90	32.70	155.80	52
Distance to Dar es Salaam (km)	UN OCHA ROSA, HDX	599.50	272.80	57.30	1091.70	52
District revenue per capita in 1966 ('000 shs)	Jensen (1968)	0.01	0.00	0.00	0.02	52
Ethnolinguistic fractionalization in 1967	Population Census	0.55	0.24	0.07	0.91	52
Centroid latitude	UN OCHA ROSA, HDX	9351933.00	320324.09	8786154.00	9850980.00	52
Centroid longitude	UN OCHA ROSA, HDX	52776.40	306742.59	-478720.59	610535.00	52
Average altitude in meters	USGS EROS	1030.00	414.00	116.40	1723.90	52
Average slope	Nunn and Puga (2012)	0.67	0.20	0.07	1.10	52
Hospital beds in 1967 (per '000.000 inhabitants)	Jensen (1968)	1.07	0.72	0.14	3.32	52
Drought in 1974 (censored z-score)	Tanzania Meteo	-0.02	0.10	-0.59	0.00	52
Drought in 1975 (censored z-score)	Tanzania Meteo	-0.04	0.14	-0.73	0.00	52
Drought in 1976 (censored z-score)	Tanzania Meteo	-0.11	0.21	-0.88	0.00	52
National identity in 1967	Prewitt et al. (1970)	0.80	0.08	0.66	0.92	48
Share urban in 1967 (% of population)	Jensen (1968)	0.03	0.05	0.00	0.24	52
Population density in 1967 (people per sq. mile)	Jensen (1968)	70.50	76.40	3.50	436.80	52
GDP per capita in 1967 ('000 shs)	Jensen (1968)	0.36	0.20	0.18	1.19	52
Tot. market agricultural production per capita in 1967 ('000 shs)	Jensen (1968)	0.08	0.07	0.01	0.34	52
Distance to Uganda (km)	FieldMaps, HDX	595.50	346.60	51.50	1245.90	52
Number of missions in 1968 (per '000 inhabitants)	Hedde-von Westernhagen and Becker (2022)	0.02	0.01	0.00	0.06	52
Total length of roads in 1968 over district area (km/km2)	Jedwab and Storeygard (2022)	0.01	0.01	0.00	0.04	52
In-migration rate 1970-2004	TNPS	0.11	0.08	0.01	0.38	52
Out-migration rate 1970-2004	TNPS	0.16	0.09	0.01	0.35	52
CCM support in 2000 and 2005 elections	Electoral data	0.59	0.12	0.14	0.83	112
Panel B. Individual level (baseline sample)						
National identity	Afrobarometer	0.89	0.23	0.00	1.00	849
Birth year	Afrobarometer	1962.30	6.48	1948.00	1971.00	849
Completed primary school	Afrobarometer	0.88	0.32	0.00	1.00	724
Believe surveyor sent by government	Afrobarometer	0.59	0.49	0.00	1.00	849
Urban	Afrobarometer	0.16	0.37	0.00	1.00	849
Cash Income Job	Afrobarometer	0.34	0.41	0.00	1.00	849
Wealth Index	Afrobarometer	-0.12	1.52	-4.03	4.16	841
Respect authority	Afrobarometer	0.21	0.31	0.00	1.00	835
Support one party rule	Afrobarometer	0.44	0.39	0.00	1.00	829
See government as parent	Afrobarometer	0.58	0.41	0.00	1.00	352
Trust government newspapers	Afrobarometer	0.08	0.21	-1.00	1.00	413
Trust government TV/radio	Afrobarometer	0.09	0.23	-1.00	1.00	422
Took action to hold government accountable	Afrobarometer	0.77	0.42	0.00	1.00	844
Trust own ethnic group	Afrobarometer	0.71	0.25	0.00	1.00	477
Trust other ethnic groups	Afrobarometer	0.64	0.27	0.00	1.00	476
Support free schooling	Afrobarometer	0.55	0.44	0.00	1.00	474
Agree women should have equal rights	Afrobarometer	0.89	0.25	0.00	1.00	477
Agree women should have equal election chances	Afrobarometer	0.91	0.23	0.00	1.00	477
Support CCM	Afrobarometer	0.94	0.24	0.00	1.00	684
Completed primary school	IPUMS	0.67	0.47	0.00	1.00	581545
Swahili Literate	IPUMS	0.71	0.45	0.00	1.00	675578
Main occupation: employed in government	TNPS	0.04	0.19	0.00	1.00	1599
Main occupation: employed in private sector	TNPS	0.02	0.15	0.00	1.00	1599
Main occupation: employed in agriculture	TNPS	0.82	0.38	0.00	1.00	1599
Main occupation: self-employed	TNPS	0.08	0.27	0.00	1.00	1599
Annual real household consumption, per adult ('000 shs)	TNPS	554.40	427.00	82.30	4072.30	1633
Annual nominal furnishings and household expenditures ('000 shs)	TNPS	56.00	144.00	0.00	3332.00	1633
Married within same ethnic group (if married, husband)	DHS	0.66	0.47	0.00	1.00	1156
Married within same ethnic group (if married, wife)	DHS	0.70	0.46	0.00	1.00	1193
Married	DHS	0.73	0.45	0.00	1.00	11552

Notes: The data includes observations from cohorts used in the main analysis (born between 1948 and 1971). The capital, Dar es Salaam (Mzizima) and the islands Zanzibar, Pemba and Mafia are excluded from all analyses. CCM support in elections is based on the election years 2000 and 2005, averaged at 1967 district level, so each district is observed twice. Data construction and sources are described in detail in Section 3, Appendix Table A.2 and Table A.3.

Table 2: The Effect of *Ujamaa* on National Identity, Main Results

	Dependent Variable: National vs. Ethnic Identity							
	no zone-cohort FE	baseline	controlling for treated cohort dummy interacted with					
			distance to capital	district revenue	ethnolinguistic fractionalization	geographical features	hospital beds	weather shocks
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Villagization × Treated Cohort	0.165**	0.226***	0.227***	0.204***	0.208**	0.214***	0.233**	0.145**
	(0.079)	(0.081)	(0.078)	(0.066)	(0.078)	(0.075)	(0.091)	(0.072)
Observations	849	849	849	849	849	849	849	849
Number of clusters	52	52	52	52	52	52	52	52
R-squared	0.121	0.139	0.141	0.144	0.140	0.148	0.142	0.143
District FE	✓	✓	✓	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓
Zone-Cohort FE		✓	✓	✓	✓	✓	✓	✓

Notes: The unit of observation is an individual Afrobarometer respondent i in district d belonging to school cohort t. All columns regress the measure of an individual's current national identity on the interaction between that individual's district-level measure of historical villagization and a dummy that indicates whether the individual is in the treated cohort, controlling for the 1967 district primary school enrollment rate interacted with the cohort dummy, survey year fixed effects, and the other controls and fixed effects indicated in the table. The un-interacted variables are included in the fixed effects in all columns. The dependent variable is from Afrobarometer rounds 3 and 4 (2005-2008), and was originally on a five-level Likert scale; it has been recoded from 0 to 1 in 1/4 increments (1 = respondent identifies only with the nation as a whole; 0 = respondent identifies only with their ethnic group). Treated cohort is a dummy that equals 1 if the respondent was born in 1960-1971. The reference group is the cohort born in 1948-1959. The villagization variable is the share of the respondent's current district's rural population that lived in registered government villages by 1978 according to the 1978 Tanzania Population Census. All controls, interacted with the individual-level cohort dummy, are at the district level and based on the 1967 Census (except weather shocks and geographical features). Geographical features include latitude, longitude, altitude and slope. Weather shocks are measured as z-scores of rainfall during the planting season in 1974, 1975 and 1976 (main period of mandatory villagization), censored at 0 to capture droughts. The weather data is based on readings from 108 weather stations which are spatially interpolated at the district level using Kriging. The villagization and outcome variables are standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. All regressions are weighted using the survey weights provided by Afrobarometer. Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: The Effect of *Ujamaa* on National Identity, Additional Controls

	Dependent Variable: National vs. Ethnic Identity									
	no zone-cohort FE	baseline	controlling for treated cohort dummy interacted with							
			national identity in 1967	share urban	population density	GDP	agricultural production	distance to Uganda border	colonial missions	roads length
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Villagization × Treated Cohort	0.165**	0.226***	0.173**	0.229***	0.232***	0.226***	0.236***	0.234***	0.241***	0.226***
	(0.079)	(0.081)	(0.075)	(0.085)	(0.084)	(0.079)	(0.076)	(0.086)	(0.069)	(0.080)
Observations	849	849	785	849	849	849	849	849	849	849
Number of clusters	52	52	48	52	52	52	52	52	52	52
R-squared	0.121	0.139	0.137	0.139	0.139	0.139	0.140	0.139	0.140	0.139
District FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Zone-Cohort FE		✓	✓	✓	✓	✓	✓	✓	✓	✓

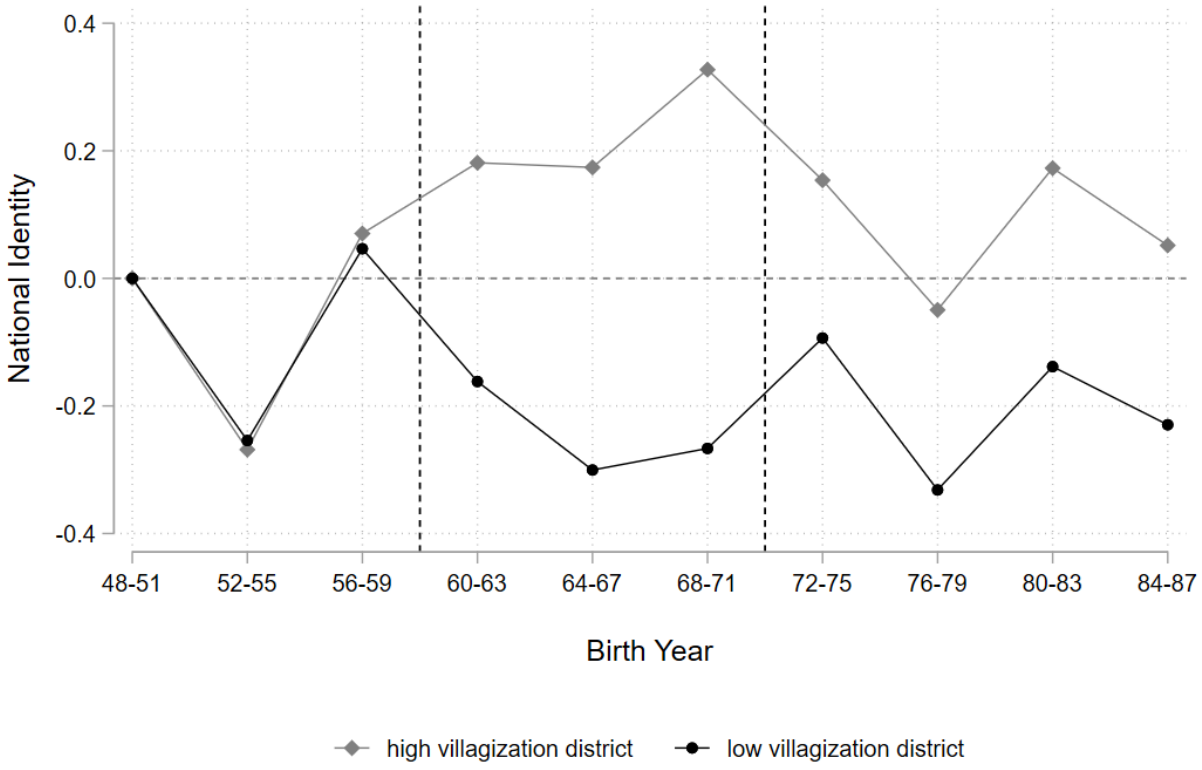
Notes: The unit of observation is an individual Afrobarometer respondent i in district d belonging to school cohort t. All columns regress the measure of an individual's current national identity on the interaction between that individual's district-level measure of historical villagization and a dummy that indicates whether the individual is in the treated cohort, controlling for the 1967 district primary school enrollment rate interacted with the cohort dummy, survey year fixed effects, and the other controls and fixed effects indicated in the table. The un-interacted variables are included in the fixed effects in all columns. The dependent variable is from Afrobarometer rounds 3 and 4 (2005-2008), and was originally on a five-level Likert scale; it has been recoded from 0 to 1 in 1/4 increments (1 = respondent identifies only with the nation as a whole; 0 = respondent identifies only with their ethnic group). Treated cohort is a dummy that equals 1 if the respondent was born in 1960-1971. The reference group is the cohort born in 1948-1959. The villagization variable is the share of the respondent's current district's rural population that lived in registered government villages by 1978 according to the 1978 Tanzania Population Census. All controls, interacted with the individual-level cohort dummy, are at the district level. The added controls in columns (3) to (10) are, respectively, national identity in 1967 (based on a nationwide survey of almost 3,000 Tanzanian secondary school students in 1967 by Prewitt et al. (1970)), urbanization, population density, GDP, the value of agricultural production (all from Jensen (1968)), distance to the border with Uganda, the number of colonial missions per capita (Hedde-von Westernhagen and Becker, 2022), and the total length of roads in the district divided by district area in 1967 (Jedwab and Storeygard, 2022). The villagization and outcome variables are standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. All regressions are weighted using the survey weights provided by Afrobarometer. Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: The Effect of *Ujamaa* on National Identity: Heterogeneity by Primary School Completion, Ethnolinguistic Fractionalization and Gender

	Dependent Variable: National Identity			
			Males	Females
	(1)	(2)	(3)	(4)
Villagization × Treated Cohort	-0.121	0.206**	0.267**	0.129
	(0.147)	(0.101)	(0.130)	(0.095)
Villagization × Completed School	-0.111			
	(0.110)			
Treated Cohort × Completed School	0.050			
	(0.478)			
Villagization × Treated Cohort × Completed School	0.331*			
	(0.188)			
Completed School	-0.049			
	(0.217)			
Treated Cohort × ELF		0.245		
		(0.260)		
Villagization × Treated Cohort × ELF		0.027		
		(0.101)		
Observations	724	849	453	394
Number of clusters	52	52	50	51
R-squared	0.171	0.142	0.227	0.185
District FE	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓
Zone-Cohort FE	✓	✓	✓	✓

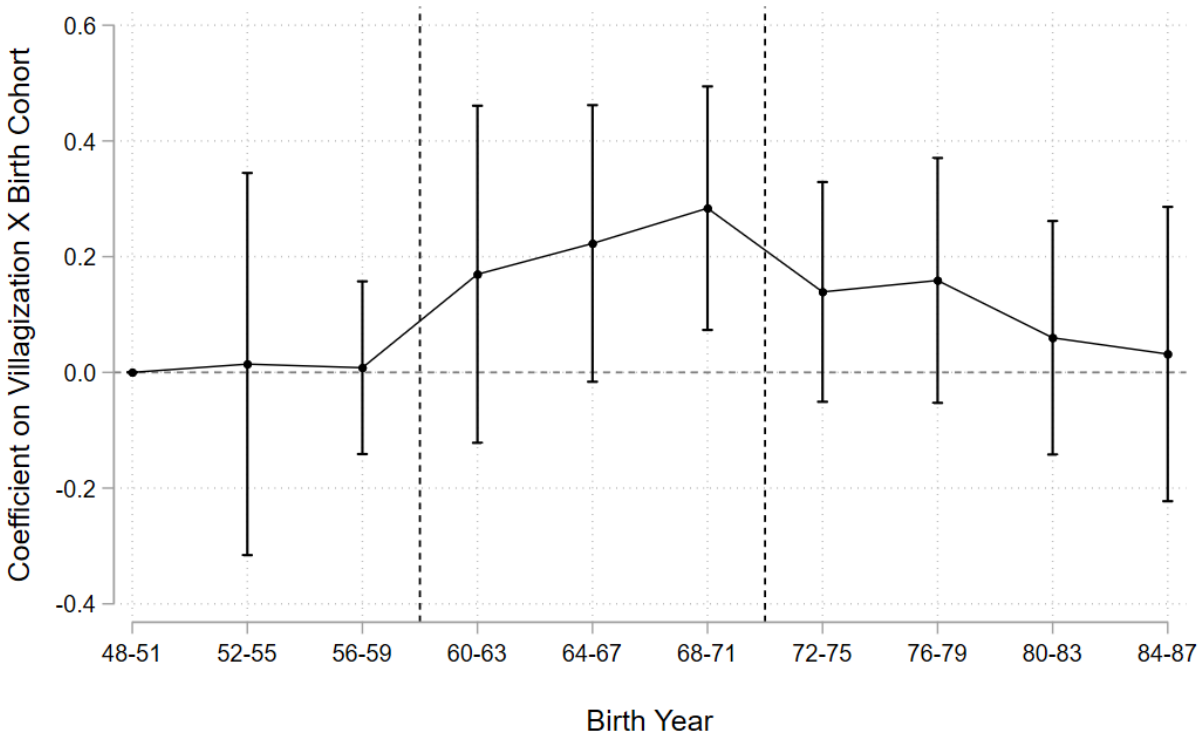
Notes: The unit of observation is an individual Afrobarometer respondent i in district d belonging to school cohort t and the outcome is a measure of an individual's current national identity. In columns (3) and (4) the sample is restricted to male and female respondents, respectively. In column (1) the dependent variable is regressed on the interaction between that respondent's district-level measure of historical villagization and a dummy that indicates whether the respondent is in the treated cohort, a dummy that indicates whether the respondent ever completed formal primary schooling, all double and triple interactions between villagization, cohort and schooling (that are not included in the fixed effects), controlling for the 1967 district primary school enrollment rate interacted with the cohort dummy and the dummy for primary school completion as well as the corresponding double interactions, survey year fixed effects, and the fixed effects indicated in the table. In column (2) the dependent variable is regressed on the interaction between the respondent's district-level measure of historical villagization, a dummy that indicates whether the respondent is in the treated cohort, and the ELF measure, controlling for the 1967 district primary school enrollment rate interacted with the cohort dummy and the ELF measure as well as the corresponding double interactions, survey year fixed effects, and the fixed effects indicated in the table. Columns (3) and (4) regress the measure of an individual's current national identity on the interaction between that individual's district-level measure of historical villagization and a dummy that indicates whether the respondent is in the treated cohort, controlling for the 1967 district primary school enrollment rate interacted with the cohort dummy, survey year fixed effects, and the fixed effects indicated in the table. The variables varying at the district- or cohort-level only are included in the fixed effects in all columns. The national identity outcome is from Afrobarometer rounds 3 and 4 (2005-2008), and was originally on a five-level Likert scale; it has been recoded from 0 to 1 in 1/4 increments (1 = respondent identifies only with the nation as a whole; 0 = respondent identifies only with their ethnic group). Treated cohort is a dummy that equals 1 if the respondent was born in 1960-1971. The reference group is the cohort born in 1948-1959. The villagization variable is the share of the respondent's current district's rural population that lived in registered government villages by 1978 according to the 1978 Tanzania Population Census. The villagization, ELF, and measure of national identity variables are standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. Regressions are weighted using the survey weights provided by Afrobarometer. Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 1: National Identity by Cohort and Villagization



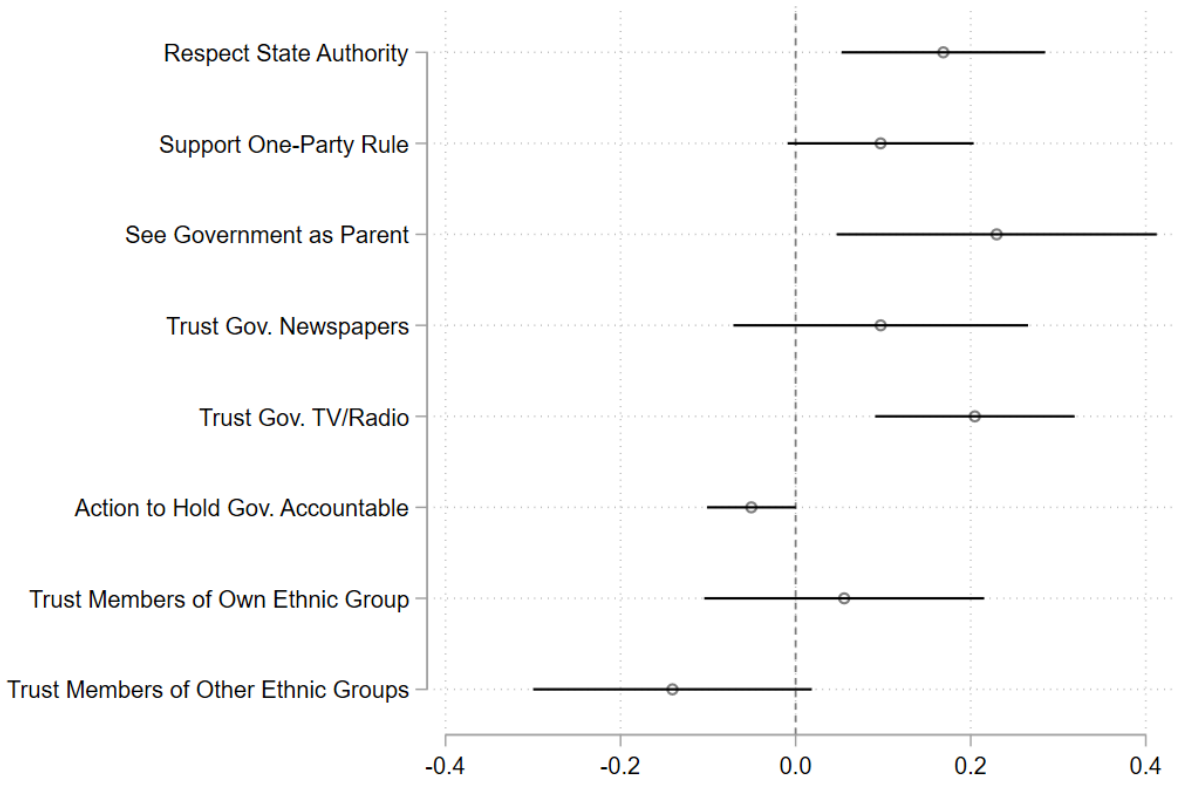
Notes: The unit of observation is an individual Afrobarometer respondent i in district d belonging to birth cohort t. Each birth cohort includes four years between 1948 and 1987 (1948-1951, 1952-1955, ...). The dashed lines indicate the first and last birth cohorts that were of 4^{th} grade age (10 years old) during the villagization period (1970-1981). The gray (black) line plots the coefficients from a regression of the measure of an individual's current national identity on birth cohort dummies among high (low) villagization districts (high = above or equal to sample median, low = below sample median). The two regressions include no fixed effects or controls. The estimates are relative to the 1948-1951 birth cohort, which is the omitted category. The dependent variable is from Afrobarometer rounds 3 and 4 (2005-2008), and was originally on a five-level Likert scale; it has been recoded from 0 to 1 in 1/4 increments (1 = respondent identifies only with the nation as a whole, 0 = respondent identifies only with their ethnic group). The villagization variable is the share of the respondent's current district's rural population that lived in registered government villages by 1978 according to the 1978 Tanzania Population Census. The outcome variable is standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. The regressions are weighted using the survey weights provided by Afrobarometer.

Figure 2: The Effect of *Ujamaa* on National Identity, All Cohorts



Notes: The unit of observation is an individual Afrobarometer respondent i in district d belonging to birth cohort t. Each birth cohort includes four years between 1948 and 1987 (1948-1951, 1952-1955, ...). The dashed lines indicate the first and last birth cohorts that were of 4th grade age (10 years old) during the villagization period (1970-1981). The thick line plots the coefficients from a regression of the measure of an individual's current national identity on the interaction between that individual's district-level measure of historical villagization, dummies that indicate whether the individual is in each of the cohorts shown, controlling for the 1967 district primary school enrollment rate interacted with the birth cohort dummies, survey year fixed effects, 1967 district fixed effects, birth cohort fixed effects and zone-cohort fixed effects. The vertical solid lines show 95% confidence intervals based on robust standard errors clustered at the district level. The estimates are relative to the 1948-1951 birth cohort, which is the omitted category. District and cohort fixed effects as well as all baseline controls (1967 district primary school enrollment rate interacted with cohort fixed effects, zone-cohort fixed effects, survey year fixed effects,) are included. The un-interacted variables are included in the fixed effects. The dependent variable is from Afrobarometer rounds 3 and 4 (2005-2008), and was originally on a five-level Likert scale; it has been recoded from 0 to 1 in 1/4 increments (1 = respondent identifies only with the nation as a whole, 0 = respondent identifies only with their ethnic group). The villagization variable is the share of the respondent's current district's rural population that lived in registered government villages by 1978 according to the 1978 Tanzania Population Census. The villagization and outcome variables are standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. The regression is weighted using the survey weights provided by Afrobarometer.

Figure 3: The Effects of *Ujamaa* on State Legitimacy and Accountability



Notes: The unit of observation is an individual Afrobarometer respondent i in district d belonging to school cohort t. The dots show the coefficients from regressions of the outcome stated on the left on the interaction between an individual's district-level measure of historical villagization and a dummy that indicates whether that individual is in the treated cohort, controlling for the 1967 district primary school enrollment rate interacted with the cohort dummy, survey year fixed effects, 1967 district fixed effects, the cohort dummy and zone-cohort fixed effects. The un-interacted variables are included in the fixed effects in all regressions. The dependent variables are from Afrobarometer rounds 3 and/or 4 (2005-2008) and are as follows (recoded from Likert scale). (1) "Choose Statement 1 or Statement 2. Do you agree or agree very strongly? Statement 1: Citizens should be more active in questioning the actions of leaders. Statement 2: In our country, citizens should show more respect for authority." The dependent variable registers agreement with Statement 2.; (2) "There are many ways to govern a country. Would you disapprove or approve of the following alternatives? Only one political party is allowed to stand for election and hold office." The dependent variable registers approval with this option.; (3) "Choose Statement 1 or Statement 2. Do you agree or agree very strongly? Statement 1: People are like children; the government should take care of them like a parent. Statement 2: Government is like an employee; the people should be the bosses who control the government." The dependent variable registers agreement with Statement 1.; (4) "How much do you trust each of the following, or haven't you heard enough about them to say: Government (relative to independent) newspapers?"; (5) "Government (relative to independent) broadcasting service (TV / radio)?""; (6) "Here is a list of actions that people sometimes take as citizens. For each of these, please tell me whether you, personally, have done any of these things during the past year" Got together with others to raise an issue"; (7) "How much do you trust each of the following types of people: "People from your own ethnic group?". (8) "People from other ethnic groups?". Questions (4), (5), (7) and (8) were asked only in Round 3, question (3) only in Round 4. Treated cohort is a dummy that equals 1 if the respondent was born in 1960-1971. The reference group is the cohort born in 1948-1959. The villagization variable is the share of the respondent's current district's rural population that lived in registered government villages by 1978 according to the 1978 Tanzania Population Census. The villagization and outcome variables, except outcome (6) which is binary, are standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. All regressions are weighted using the survey weights provided by Afrobarometer. Standard errors are clustered at the district level. The horizontal bars show 90% confidence intervals.

ONLINE APPENDIX

Appendix A. Correlates of Villagization

Our empirical strategy exploits variation in villagization intensity across space, as well as variation in age cohorts that were differentially exposed to the *Ujamaa* policy. Variation in villagization intensity across space can largely be explained by the fact that its implementation was left to local officials, who enforced the policy with different degrees of conviction (McHenry, 1979). This variation may not be random – nor does it need to be for our empirical strategy to be valid, as we explain in Section 4. Nevertheless, we control for the most important potential correlates of villagization intensity, as identified in historical accounts. As discussed in turn below, these include: pre-*Ujamaa* primary school enrollment rate, distance to capital (to capture central government influence), district revenues (to capture local government capacity), ethnolinguistic fractionalization, geographic characteristics, availability of public health infrastructure, weather shocks, pre-policy feelings of national identity, the degree of urbanization and population density, GDP and agricultural production (to capture economic development), distance to Uganda (to capture potential exposure to the Uganda-Tanzania conflict), and the presence of colonial missions and roads (Ergas, 1980; Hydén, 1980). Table A.12 documents the pairwise correlations of the relevant variables with villagization (conditional on zone fixed effects, as in our baseline specification). We use newly digitized district-level variables from official government statistics (Jensen, 1968) based on the 1967 population census, the last census before the villagization policy was implemented, and various other sources (see Table A.2 for variable definitions and sources).[45]

Primary School Enrollment

The data shows a positive correlation between our measure of villagization (share of the rural district population living in planned villages in 1978) and the pre-policy primary school enrollment rate. This positive correlation can be explained by the fact that the government was more successful in implementing the policy in areas that already had existing (schooling) infrastructure. The government used access to schools and other public services such as dispensaries and piped water supply to encourage people to move to *Ujamaa* villages (McHenry, 1979; Boesen et al., 1977; Scott, 1998). Settlements that already featured some

[45] We made small corrections to the variables in Jensen (1968) in the process of digitization to address inconsistencies in the original data and clear typos. These corrections are outlined in the replication package of the paper in detail. Briefly, first, we exclude Mzizima district, the urban part of which included Dar es Salaam in 1967. This is consistent with our sample in the paper. We also adjust the amount of district revenue in proportion to population for two pairs of districts for which revenue data had been aggregated (Iringa/Mufindi and Mbeya/Mbozi). We also assume that Moshi Urban is part of Kilimanjaro district in 1967 and fix a clear typo in the reported population of Lushoto district.

of this infrastructure thus made for easier targets. In any case, primary school enrollment levels were low (7.1%) prior to villagization.

Central Government Influence

We next consider the correlation between the distance to the capital city, Dar es Salaam, as a measure of central government influence, and the intensity of villagization.[46] Given limited control of the hinterland by the state (Herbst, 2014), we might expect that the policy was particularly targeted to the remote parts of the country to bring those areas into the state's orbit. Scott (1998) also notes that sites for new village locations were chosen not by any economic logic, but by finding "blank spots" on the map where settlers might be relocated, citing Boesen et al. (1977). However, we find no significant relationship between distance to Dar es Salaam and villagization intensity. This can be explained by our use of variation at the district level (as opposed to within-district variation) and inclusion of zone fixed effects.

Local Government Capacity

Given that district officials played a key role in *Ujamaa* implementation (McHenry, 1979), we might expect districts with higher local government capacity to implement the policy more intensely – both in terms of resettling the local population and implementing the concurrent education reform. The central government did not have sufficient funds to build all of the new schools required to meet the universal primary education targets laid out in the policy, and thus promoted "self-help" efforts by district and village officials (Stabler, 1979). We proxy for local state capacity using district government revenue per capita. However, we find a close to zero correlation between pre-*Ujamaa* government revenue and villagization.

Ethnolinguistic Fractionalization

We next look at the relationship between villagization and ethnolinguistic fractionalization (ELF) in 1967. This reflects the idea that ethnic diversity may have hindered the population's willingness to comply with the villagization policy, which involved potentially co-inhabiting villages with members of different ethnic groups. Consistent with this argument, we find that the intensity of villagization was slightly lower in districts with high pre-*Ujamaa* ELF, albeit the correlation is statistically insignificant.

[46] Dar es Salaam was not only the seat of government in the pre-policy era but also the country's main administrative and commercial center. In 1973, the government made the decision to move the national capital to Dodoma within 10 years (Hayuma, 1980).

Geography

Historical accounts indicate that the most important criterion for the location of Ujamaa villages was convenient access (Hydén, 1980). This implies that geographic characteristics such as topology or climatic zones may have also influenced the degree of villagization. For example, more rugged or remote terrain may have been harder to consolidate. We find no significant correlations between several geographic characteristics (altitude, slope, latitude and longitude) and villagization. This can be explained by the fact that we exploit cross-district rather than within-district variation in villagization and condition on zone fixed effects.

Public Health Infrastructure

We test whether there is a relationship between the intensity of villagization and existing public health infrastructure, proxied by the number of hospital beds per hundred thousand inhabitants in 1967. We find no significant correlation.

Weather Shocks

Several regions in Tanzania experienced severe droughts during the height of the resettlement period (1974–1976). Hydén (1980) argues that from the point of view of the state, the drought served as a "blessing in disguise [that] facilitated the movement of people" (p. 146). The drought resulted in a major food shortage (Ergas, 1980). The government also provisioned drought relief to incentivize people to move into planned villages. Indeed, in a study on the motives behind migration to villages, Bakula (1971, in Sitari (1983)) finds that 80% of migrants say the main reason they moved was hope of increased monetary income. Thus, we check whether unexpected weather shocks in the form of droughts affected villagization. In line with the historical accounts, we find a positive correlation between drought occurrence in 1976 and villagization.[47]

Pre-policy Feelings of National Identity

Figure A.8 depicts the association between villagization in 1978 and feelings of national identity in 1967 (i.e, prior to the *Ujamaa* policy) at the region level.[48] The data on

[47] We measure droughts using the standard deviation in yearly rainfall from a district's long-term mean rainfall (1960–2010) during the planting season (March to May), censored at 0 (following Dell et al. (2014)). Rainfall is calculated based on readings from 108 weather stations across Tanzania and aggregated to the district level using Kriging (spatial interpolation).

[48] The 1967 survey by Prewitt et al. (1970) does not contain district-level identifiers. We therefore linked the region-level survey data to the districts in our sample, which are each assigned to a region. We thank Ted Miguel and Tina Green for help with accessing the data.

national identity in 1967 is based on a nationwide survey of almost 3,000 secondary students (Prewitt et al., 1970) and is similar in wording to the measure of national identity from the contemporary Afrobarometer data used in the paper. We find a weakly negative relationship between villagization and pre-*Ujamaa* feelings of national identity. This suggests that, in line with the government's stated intention of building national identity using villagization, the policy was aimed at regions that had lower initial levels of identification with the nation.

Urbanization and Population Density

We find no significant correlation between villagization intensity and the pre-*Ujamaa* degree of urbanization (1 – share of the rural district population in 1967). Urbanization levels were very low (3% of the country's population according to the 1967 population census) prior to villagization. We find a positive but statistically insignificant relationship between villagization intensity and pre-*Ujamaa* population density.

Economic Development

Scott (1998) notes that some regions (in particular, West Lake and Kilimanjaro) saw very little impact of *Ujamaa*. These regions were spared for three reasons: (1) farmers were already living in populous villages; (2) their undisturbed productivity of cash crops was vital for state revenue; and (3) groups residing in these areas were over-represented in the bureaucratic elite. Ergas (1980) and Sitari (1983) echo this, noting the poorest response to villagization in densely settled, prosperous territories. These areas were home to prosperous farmers, and were places where cash crops were produced in large amounts. Indeed, according to Barker (1974), the number and population of *Ujamaa* villages appears to vary inversely with importance of cash crops in region. We find positive but insignificant correlations between agricultural production and per capita GDP before *Ujamaa*, conditional on zone fixed effects.

Other Variables

Villagization occurred during a period that spanned war with Uganda. In response to Idi Amin's annexation of Kagera region in northwestern Tanzania in November 1978, Nyerere launched a counter-offensive that ultimately removed Amin from power in April 1979 (Roberts, 2014). Although the official loss to Tanzania in terms of human life was relatively low (373 soldiers killed, of whom 96 died due to enemy fire), the ultimate cost of the intervention and occupation of Uganda was estimated to be more than $500 million (Roberts, 2014). We find that villagization was stronger the farther away from the border with Uganda conditional on zone fixed effects, although the correlation coefficient is insignificant.

Finally, the number of historical missions per capita and the length of roads divided by district area are not significantly correlated with villagization.

Appendix B. Migration

Here we conduct two exercises to show that selective migration is unlikely to explain most of the positive relationship between the *Ujamaa* policy and national identity we observe. Recall that only inter-district migration potentially threatens the validity of our estimates.

The Afrobarometer data do not contain information on respondents' migration history. However, we can calculate inter-district migration rates, during the period after *Ujamaa* and before the Afrobarometer survey took place, using the Tanzania National Panel Survey 2008 data.[49] That is, we can calculate the in- and out-migration rate for each origin-/destination-district-combination. We use the migration matrix constructed this way for the robustness checks related to selective migration.

As a first exercise, in Table A.6 column (5) we drop 14 out of 52 districts that are in the top deciles in terms of either in- or out-migration rates. If (selective) migration were driving our results, we should see a clear drop in our main coefficient when we exclude such high-migration districts. Reassuringly, even though we lose some statistical power due to the reduced sample size, the magnitude of the coefficient with this reduced sample is similar to the coefficient for our baseline specification with the full sample.

As a second exercise, we quantify the potential measurement error stemming from selective migration to provide bounds on our main coefficient under different assumptions about the extent of selective migration. We begin by noting that migration may or may not lead to measurement error in a respondent's villagization measure: for example, if a respondent in the treated cohort is at the time of survey (2005-2008) in a district with a villagization measure of 0.5, but was in a district with a villagization measure of 1 during the villagization period (1970–1981), then the measurement error in their treatment amounts to -0.5. However, if they migrated from a district with a villagization measure of 1 to another district with a villagization measure of 1, the measurement error is 0 and does not bias our estimates.

The expected measurement error in the villagization measure among respondents in a district thus depends on two factors: 1) the difference in villagization between the receiving district and the sending district, and 2) the migration rates between each district-pair. To estimate the expected measurement error in our sample, we proceed as follows. We first calculate the measurement error for each district-pair (receiving and sending) as the product

[49] The Tanzania National Panel Survey data contains variables on each respondent's current district, previous district, and year of migration (if migrated). We calculate migration rates based on migration during 1970 and 2004 because we are concerned about migration caused by the *Ujamaa* policy in 1970-1981 before the Afrobarometer survey (2005-2008). We assume that if the respondent migrated between 1970 and 2004, the origin district is the district where they resided at the beginning of the villagization period.

of the difference between the two districts' villagization measures and the rate of migration from the sending to the receiving district:[50]

$$error_{ds} = (V_d - V_s) \cdot m_{ds} \tag{A 1}$$

where V_d and V_s denote villagization in receiving and sending districts d and s, and $m_{d,s}$ denotes the migration rate from district s to district d. Migration from districts with higher villagization thus leads to negative measurement error in the villagization measure, while migration from districts with lower villagization leads to positive measurement error. The larger the migration rate from a district, the larger the potential measurement error due to migration from that district.

Next, we make a strong assumption that would lead to selective migration working against our main finding: assume that residents in the current district with the national identity outcome equal to y (ranging from 0 to 1) came from a given district with lower villagization with probability $m_{ds} \cdot y_{id}$ and from a given district with higher villagization with probability $m_{ds} \cdot (1 - y_{id})$. That is, residents with higher (lower) national identity were more likely to migrate from districts with lower (higher) villagization. We adjust the villagization measures of the respondents in our main sample by the resulting expected error after summing over sending districts:[51]

$$\tilde{V}_{id} = V_d - \sum_{s=1}^{S} error_{ds} \cdot \mathbb{1}\left[error_{ds} > 0\right] \cdot y_{id} - \sum_{s=1}^{S} error_{ds} \cdot \mathbb{1}\left[error_{ds} < 0\right] \cdot (1 - y_{id}) \tag{A 2}$$

where $\tilde{V}_{id}$ denotes individual i in district d's "error-adjusted" villagization measure and y_{id} denotes national identity (ranging from 0 to 1).

Finally, we re-estimate our main specification using this adjusted villagization measure:

$$y_{idzt} = \gamma(\tilde{V}_{id} \cdot treatedcohort_t) + (\tilde{\mathbf{X}}'_{dz} \cdot treatedcohort_t)\mathbf{\Gamma} + \alpha_{dz} + \delta_{zt} + \epsilon_{idzt} \tag{A 3}$$

The coefficient estimate of γ then provides a lower bound to our baseline coefficient estimate of β in the presence of selective migration to such an extent. Table A.6 column (6) displays the result: the estimated coefficient equals 0.144 ($p < 0.05$). As a reminder, our baseline coefficient estimate equals 0.226. That is, at most 36% of our baseline coefficient estimate can be explained by such selective migration. However, this case is very unlikely. It effectively assumes that the individuals that migrated to low-villagization districts identify

[50] For example, assume that district A has a villagization measure of 0.8 and district B has a villagization measure of 0.6. Further assume that 5% of current residents in district A migrated from district B. Then, the expected measurement error in the villagization measure among the current residents of district A stemming from migration from district B equals $(0.8 - 0.6) \cdot 0.05 = 0.01$.

[51] We adjust the 1967 district primary school enrollment rate control variable analogously.

with their ethnic group, and that the individuals that migrated to high-villagization districts identify with Tanzania as a whole. Alternatively, assume that only some of those with low national identity systematically migrated to low-villagization districts, whereas there was no selective migration among those with high national identity (that is, they did not systematically go to districts with higher villagization). In this case, the adjusted coefficient estimate, reported in Table A.6 column (7), equals 0.164 ($p < 0.05$). That is, under this more realistic assumption, at most 27% of our main result can be explained by selective migration of this nature. In sum, under reasonable assumptions, it is unlikely that selective migration explains a large part of our main result.

Appendix C. Additional Robustness Checks

Here we discuss robustness checks related to alternative choices of regression weights, sample and functional form, and address concerns related to social desirability bias.

Sample weighting. In our preferred specification, we use survey weights from Afrobarometer to make our sample nationally representative. Table A.6 column (2) shows the unweighted coefficients. The estimates from the weighted and unweighted regressions are qualitatively similar in size and significance.

Sample choice. Since villagization was a rural policy, those living in rural areas during the *Ujamaa* period should be affected by the treatment (the "compliers"). Table A.6 column (4) reports our coefficient estimates for the sample of respondents living in a rural area at the time of the survey. As expected, we find that the effect of *Ujamaa* on national identity is larger for this sub-sample (as they are more likely to be compliers).

Functional form. In Table A.6 columns (10) and (11), we report results for ordered probit and ordered logit specifications. The results are consistent with our baseline result, which is based on the linear probability model. See the table notes of Table A.6 for details.

Social desirability bias. Our main outcome measure – whether the respondent reports identifying more with their ethnic group or with the nation as a whole – may suffer from social desirability bias. The treated respondents were taught the importance of the nation and a sense of duty towards the state as part of the *Ujamaa* policy. Hence, treated respondents may believe that they are expected to respond in a certain way because they perceive the surveyor to be a representative of the national government. This could bias our coefficient of interest upwards. Afrobarometer invests considerable effort in mitigating such concerns by hiring independent surveyors unaffiliated with the government and from outside the survey districts. Nevertheless, almost 60% of the respondents in our sample reported believing that the surveyors were sent by a government institution. To account for this, we directly control for whether a respondent believes that the surveyor was sent by a government institution.

Our estimates remain unchanged when including this control variable (see Table A.6 column (3)). Social desirability bias is unlikely to explain our results.

Appendix D. Intermarriage

We here examine the effect of the *Ujamaa* policy on ethnic intermarriage. We follow a similar empirical approach as in our main specification but construct a new outcome variable using data from the DHS in 1991 and 1996.[52] We construct an outcome variable that equals 1 if a respondent's partner belongs to the same ethnic group as the respondent. The treatment variable is based on the husband or wife's district and birth year.

Table A.4 presents the results. Using the husband's cohort (column (1)), we find that a one standard deviation increase in villagization is associated with a 5.7 percentage points (0.121 standard deviations) decrease in the likelihood of the wife sharing the same ethnic group for husbands from cohorts of primary schooling age during the villagization period, compared to pre-villagization cohorts. The effect is statistically significant at the 10% level. In other words, the *Ujamaa* policy increased the rate of inter-ethnic marriages. This is consistent with *Ujamaa* leading to more inter-ethnic interactions and increased personalized trust.

One potential concern with this exercise is that our conditioning variable – whether a respondent is married – could be endogenous to our treatment. However, as column (3) in Table A.4 shows, our treatment has no large or statistically significant effect on the probability that a given individual in the DHS rounds we study is married.

Column (2) shows the coefficient of the same exercise based on the wife's instead of the husband's cohort. Although the effect of *Ujamaa* on intermarriage goes in the same direction, the coefficient is smaller and statistically insignificant at conventional levels. We identify three main reasons for a discrepancy between the results based on the husband's and wife's cohort: first, wives are on average seven years younger than their husbands. Thus, if a wife is treated, her husband tends to be in the older cohort. To corroborate this explanation, columns (4) and (5) interact the *Ujamaa* treatment with the age difference between husband and wife. Consistent with our explanation, we find that the results on intermarriage based on wife's birth year are largest and significant for couples with a small age difference. A second explanation reflects gender norms. Women's autonomy in marriage choice tends to be constrained in Tanzania, so we would expect results on intermarriage to be weaker when using the wife's birth cohort.[53] Thirdly, girls may have been more weakly

[52] These two survey waves record the ethnic group of each respondent and identifiers that allow us to link couples. Subsequent waves did not ask about ethnicity. We include all married couples in the sample, except those where both spouses' ethnicities were grouped as 'Foreign' or 'Other' as our outcome is not clearly defined for these couples.

[53] Recent data from Tanzania indicates that 36 percent of women marry before the age of 18, and that 30 percent have limited say in choosing their life partners (Green et al., 2023).

affected by the treatment than boys due to lower rates of school attendance, which reflect gendered patterns of household labor (Chamie, 1983). The political education curriculum also coincided with the onset of puberty, which has been shown to further constrain girls' school attendance given insufficient menstrual hygiene management resources and practices (Benshaul-Tolonen et al., 2020). Women may thus have been exposed to *Ujamaa* at lower rates and the treatment effects based on women's cohorts would be weaker than those based on men's cohorts. Consistent with this explanation, we also find a smaller effect of *Ujamaa* on national identity for women than for men, as Table 4 columns (3) and (4) show.

Appendix E. The End of *Ujamaa*

A question is whether the external developments that explain the end of *Ujamaa*, including a global economic crisis and the invasion by Uganda (see the Background section for a summary), had cohort-specific effects that could explain our main result. We address this question in two ways.

First, we test whether important factors leading to the end of *Ujamaa*, including economic factors and the war with Uganda, could explain our results. In brief, we do not find any empirical support for such an alternative explanation. Economic factors are unlikely to explain our cohort-specific results given the results in Table A.9 discussed in the paper. In addition, Table 3 column (8) shows that our estimate of the effect of *Ujamaa* on national identity is unaffected by the inclusion of distance to the Ugandan border interacted with the cohort dummy, which proxies for cohort-specific exposure to the Ugandan invasion.

Second, to assess whether the end of *Ujamaa* had cohort-specific effects, we exploit variation across cohorts that were all not exposed to our treatment directly and are thus comparable, but that differ in their exposure to the end of *Ujamaa*. Specifically, we compare the cohort born in 1948–1959 (i.e., those too old to attend primary school during *Ujamaa* but who experienced the end of the villagization program) to a younger "control" cohort born in 1976–1987 (i.e., cohorts entering primary school only after the end of villagization) in more or less intensively villagized districts.[54] Table A.6 column (12) shows that we do not find any economically or statistically significant differences in national identity between these cohorts, and thus conclude that the end of *Ujamaa* did not have cohort-specific effects.

[54] Note that one could also compare the outcomes of the treated cohorts (i.e., those in highly villagized districts during the villagization program) with younger cohorts (i.e., cohorts entering primary schooling age after the end of the villagization program). However, this comparison would not allow for separate identification of the effect of exposure to the end of *Ujamaa* and *Ujamaa* itself.

Appendix F. Additional Tables

Table A.1: Timeline

Year	Event	Details
1961	Independence	Tanganyika (Tanzania Mainland) becomes independent with Julius Nyerere as prime minister.
1963	One-party state declared	Nyerere makes TANU the sole national party in the stated interest of the whole nation and economic development (Nyerere, 1963).
1967	*Arusha Declaration*	Declaration formed basis for national ideology; outlined plans for nation-building policies and *Ujamaa* villages.
1967	*Education for Self-Reliance*	Formed basis of *Ujamaa* education reforms.
1967	*Education Circular No. 2*	Swahili made language of instruction in public primary schools.
1967	1967 Population Census	First population census after independence conducted; administrative districts defined.
1968	Roll-out of new education policies	Political education added as a new subject to primary school curriculum (grades 4 to 7); new syllabi and textbooks emphasizing national identity introduced in social studies, history, geography and civics.
1969	**Presidential Circular No. 1**	**Mandatory villagization begins.**
1975	*Villages & Ujamaa Villages Act*	Villagization formalized in the law.
1978	1978 Population Census	First census after villagization was declared concluded.
1982	**Repeal of *Villages* & Ujamaa *Villages Act***	**Villagization period ends.**
1985	Nyerere steps down	Ali Hassan Mwinyi replaces Nyerere as prime minister, abandons many of Nyerere's policies and enacts reforms.
1992	Civics reintroduced	Civics replaces Political Education in public school curriculum.

Notes: This table provides a timeline of activities by Nyerere and the Tanzanian state after Independence and during the *Ujamaa* period. See sources described in Section 2 for additional details.

Table A.2: District-level Variables: Sources and Definitions

Variable	Source	Details and Coding
Villagization 1978	Population Census	Total rural district population living in registered villages (as opposed to unregistered villages, scattered and migratory population) divided by total district population, excluding the institutional population, as classified by the 1978 population census.
Primary school enrollment rate in 1967	Jensen (1968)	Number of children enrolled in primary schools per one thousand inhabitants according to official government statistics (Jensen, 1968).
Distance to Dar es Salaam	UN OCHA ROSA, HDX	Distance of the district centroid to Tanzania's capital city in kilometers.
District revenue per capita in 1966	Jensen (1968)	Total district revenue per capita in thousands of Tanzanian Shillings according to official government statistics (Jensen, 1968).
Ethnolinguistic fractionalization in 1967	Population Census	Calculated as 1 minus the Herfindahl concentration index of ethnolinguistic group shares in each district given by the 1967 population census data.
Centroid latitude / longitude	UN OCHA ROSA, HDX	Latitude / longitude of the district centroid in degrees.
Average altitude	USGS EROS	Mean district altitude in meters, calculated from gridded topology data.
Average slope	Nunn and Puga (2012)	Mean slope, calculated from gridded topology data (Nunn and Puga, 2012).
Hospital beds in 1967	Jensen (1968)	Number of hospital beds per 100,000 inhabitants according to official government statistics (Jensen, 1968).
Drought in 1974-1976	Tanzania Meteo	Average amount of yearly rainfall per district during the planting season (March to May), in terms of standard deviations from the long-run district mean (1960–2010), censored at 0. Data constructed from readings of 108 weather stations using Kriging (spatial interpolation) (Osafo-Kwaako, 2012).
National identity in 1967	Prewitt et al. (1970)	Based on a nationwide survey of almost 3,000 Tanzanian secondary school students in 1967 (Prewitt et al., 1970). The variable is coded as 1 if a respondent feels that the nation is more important than the tribe and 0 otherwise.
Share urban in 1967	Jensen (1968)	Share of total district population living in urban wards according to official government statistics (Jensen, 1968).
Population density in 1967	Jensen (1968)	Total district population in 1967, divided by 1967 district area (square miles) (Jensen, 1968).
GDP per capita in 1967	Jensen (1968)	Total district GDP in 1967, divided by district population in 1967 (Jensen, 1968).
Tot. market agricultural production per capita in 1967	Jensen (1968)	Total district market agricultural production in 1967, divided by district population in 1967 (Jensen, 1968).
Distance to Uganda	FieldMaps, HDX	Distance to the closest border with Uganda, in km.
Number of missions per capita in 1967	Hedde-von Westernhagen and Becker (2022)	Number of colonial missions in the district in 1967, divided by district population in 1967 (Hedde-von Westernhagen and Becker, 2022).
Total length of roads over district area in 1967	Jedwab and Storeygard (2022)	Total length of roads in the district in 1967 (km), divided by district area (square km) (Jedwab and Storeygard, 2022).
In- / out- migration rate 1970-2004	TNPS	Fraction of district population that moved from/to a different district between 1970 and 2004. Calculated from individual-level data on household heads in the TNPS Round 1 (2008).
CCM support in elections	Electoral data	Product of turnout and CCM presidential vote share from the 2000 and 2005 presidential elections in Tanzania (Carlitz, 2017), aggregated at the level of 1967 districts averaging over constituencies.

Notes: This table provides the key district-level variables, their sources and coding definitions used in our empirical analysis. There are 52 districts in our main sample. Section 3 provides additional discussion.

Table A.3: Individual-level Variables: Sources and Definitions

Variable	**Source**	**Details and Coding**
National identity	Afrobarometer	"Let us suppose that you had to choose between being a [nationality] and being a [respondent's ethnic group]. Which of the following statements best expresses your feelings?" 0 = ethnic only, 1/4 = ethnic more than national, 1/2 = ethnic and national equal, 3/4 = national more than ethnic, 1 = national only.
School cohort	Afrobarometer	Constructed from birth year of respondent plus 10 years (typical age of 4th grader). "Treated cohort" = born 1960-1971, "control cohort" = born 1948-1959.
Completed school	Afrobarometer / IPUMS	"What is the highest level of education you have completed?" 1 = completed formal primary schooling, 0 = completed no formal schooling.
Believe surveyor sent by government	Afrobarometer	"Who do you think sent us to do this interview?" 1 = government or other public institution, 0 = otherwise.
Urban	Afrobarometer	"Urban or Rural Primary Sampling Unit" 0 = rural, 1 = urban.
Cash income job	Afrobarometer	"Do you have a job that pays a cash income? Is it full-time or part-time?" 0 = no, 1/2 = yes, part time, 1 = yes, full time.
Wealth index	Afrobarometer	First principal component of the questions asking about the frequency of the household not having enough food, clean water, medicines, cooking fuel, cash income, and about ownership of a radio, a tv and a motorbike.
Respect authority	Afrobarometer	"Let's talk for a moment about the kind of society we would like to have in this country. Which of the following statements is closest to your view? Choose Statement A or Statement B. A: As citizens, we should be more active in questioning the actions of our leaders. B: In our country these days, we should show more respect for authority." 0 = agree very strongly with A, 1/4 = agree with A, 1/2 = agree with neither, 3/4 = agree with B, 1 = agree very strongly with B.
Support one-party rule	Afrobarometer	"There are many ways to govern a country. Would you disapprove or approve of the following alternatives? Only one political party is allowed to stand for election and hold office." 0 = strongly disapprove, 1/4 = disapprove, 2/4 = neither approve nor disapprove, 3/4 = approve, 1 = strongly approve.
See government as parent	Afrobarometer	"Lets talk for a moment about the kind of society you would like to have in this country. Which of the following statements is closest to your view? Choose Statement 1 or Statement 2.Do you agree or agree very strongly? Statement 1: People are like children; the government should take care of them like a parent. Statement 2: Government is like an employee; the people should be the bosses who control the government."0 = agree very strongly with 2, 1/4 = agree with 2, 1/2 = agree with neither, 3/4= agree with 1, 1= agree very strongly with 1.
Trust government newspapers / TV / radio	Afrobarometer	"How much do you trust each of the following, or haven't you heard enough about them to say: Government / independent newspapers / TV / radio?" 0 = not at all, 1/3 = just a little, 2/3 = somewhat, 1 = a lot. For each medium, the reported trust for the independent medium is subtracted from the reported trust for the Government one.
Took action to hold government accountable	Afrobarometer	"Here is a list of actions that people sometimes take as citizens. For each of these, please tell me whether you, personally, have done any of these things during the past year. If not, would you do this if you had the chance: Got together with others to raise an issue? Attended a demonstration or protest march? " 0 = if has never done any of these things, 1 = has done at least one of these things at least once.
Trust own / other ethnic group(s)	Afrobarometer	"How much do you trust each of the following types of people: People from your own ethnic group? / Tanzanians from other ethnic groups?" 0 = not at all, 1/3 = Just a little, 2/3 = I trust them somewhat, 1 = I trust them a lot
Support free schooling	Afrobarometer	"Which of the following statements is closest to your view? A: It is better to have free schooling for our children, even if the quality of education is low. B: It is better to raise educational standards, even if we have to pay school fees." 1 = agree very strongly with A, 3/4 = agree with A, 2/4 = agree with neither, 1/4 = agree with B, 0 = agree very strongly with B.
Agree women should have equal rights	Afrobarometer	"Which of the following statements is closest to your view? A: In our country, women should have equal rights and receive the same treatment as men do. B: Women have always been subject to traditional laws and customs, and should remain so." 1 = agree very strongly with A, 3/4 = agree with A, 2/4 = agree with neither, 1/4 = agree with B, 0 = agree very strongly with B.

Continued on next page

Table A.3 – continued from previous page

Variable	Source	Details and Coding
Agree women should have equal election chances	Afrobarometer	"Which of the following statements is closest to your view? A: Women should have the same chance of being elected to political office as men. B: Men make better political leaders than women, and should be elected rather than women." 1 = agree very strongly with A, 3/4 = agree with A, 2/4 = agree with neither, 1/4 = agree with B, 0 = agree very strongly with B.
Support CCM	Afrobarometer	"Do you feel close to any particular political party?" 1 = yes, Chama cha Mapinduzi (CCM), 0 = yes, party different than CCM.
Swahili Literate	IPUMS	"Does individual know how to read and write in Kiswahili?" 1 = yes, 0 = no.
Main occupation	TNPS	Dummies for different occupations that equal 1 if household head's occupation is in government (including parastatal) / private sector / agriculture / self-employment, and 0 otherwise.
Annual real household consumption, per adult	TNPS	Total annual real household consumption per adult equivalent.
Annual nominal furnishings and household expenditures	TNPS	Total annual nominal household expenditures on furnishings and household items.
Married within ethnic group	DHS	= 1 if respondent's spouse shares self-reported ethnic group, 0 otherwise. Defined for married couples only.
Married	DHS	= 1 if respondent is currently married, 0 otherwise.

Notes: This table provides the key individual-level variables, their sources and coding definitions used in our empirical analysis. Section 3 provides additional discussion.

Table A.4: The Effect of *Ujamaa* on Intermarriage

Dep. Variable:	Married Same Ethnicity		Married	Married Same Ethnicity	
	Husband Cohort	Wife Cohort		Husband Cohort	Wife Cohort
	(1)	(2)	(3)	(4)	(5)
Villagization × Treated Cohort	-0.057*	-0.024	0.007	-0.032	-0.148**
	(0.031)	(0.026)	(0.010)	(0.049)	(0.069)
Villagization × Age Difference				0.004	-0.009*
				(0.004)	(0.005)
Treated Cohort × Age Difference				0.024	0.004
				(0.018)	(0.011)
Villagization × Treated Cohort × Age Difference				0.000	0.013**
				(0.007)	(0.005)
Age Difference				-0.007*	-0.000
				(0.004)	(0.004)
Observations	1,156	1,193	11,552	1,156	1,193
Number of clusters	52	52	52	52	52
R-squared	0.109	0.082	0.035	0.113	0.087
District FE	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓
Zone-Cohort FE	✓	✓	✓	✓	✓
Sample	Married Couples	Married Couples	Individuals	Married Couples	Married Couples

Notes: The unit of observation in columns (1), (2), (4) and (5) is a married couple i of DHS respondents in district d. The data includes the districts in the main sample as in Table 2. School cohort t is assigned based on the husband's birth year in columns (1) and (4) and the wife's birth year in columns (2) and (5). In these columns, the dependent variable is a dummy for whether the spouses are of the same ethnic group. The sample excludes couples where both spouses' ethnicities were grouped as 'Foreign' or 'Other' as our outcome is not clearly defined for them. In column (3) the unit of observation is a DHS respondent i in district d and school cohort t and the dependent variable is a dummy for whether the respondent is currently married. In columns (1) to (3) the dependent variable is regressed on the interaction between the observation's district-level measure of historical villagization and a dummy that indicates whether the observation is in the treated cohort, controlling for the 1967 district primary school enrollment rate interacted with the cohort dummy, survey year fixed effects, and the fixed effects indicated in the table. In columns (4) and (5) the dependent variable is regressed on the interaction between that observation's district-level measure of historical villagization and a dummy that indicates whether the observation is in the treated cohort, and the absolute value of the age difference between the husband and the wife, all double and triple interactions between villagization, cohort and age difference (that are not included in the fixed effects), controlling for the 1967 district primary school enrollment rate interacted with the cohort dummy and age difference as well as the corresponding double interactions, survey year fixed effects, and the fixed effects indicated in the table. The variables varying at the district- or cohort-level only are included in the fixed effects in all columns. Treated cohort is a dummy that equals 1 if the respondent was born in 1960-1971. The reference group is the cohort born in 1948-1959. The villagization variable is the share of the observation's current district's rural population that lived in registered government villages by 1978 according to the 1978 Tanzania Population Census. The villagization measure is standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5: The Effect of *Ujamaa* on National Identity, All Cohorts

	Dependent Variable: National vs. Ethnic Identity							
	no zone-cohort FE	baseline	controlling for treated cohort dummy interacted with					
			distance to capital	district revenue	ethnolinguistic fractionalization	geographical features	hospital beds	weather shocks
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Villagization × Birth Year 1952-1955	-0.006	0.014	0.008	-0.047	-0.013	-0.004	-0.083	-0.064
	(0.158)	(0.168)	(0.172)	(0.172)	(0.169)	(0.164)	(0.227)	(0.184)
Villagization × Birth Year 1956-1959	-0.034	0.008	-0.001	0.011	-0.037	-0.006	0.006	-0.062
	(0.075)	(0.076)	(0.085)	(0.063)	(0.089)	(0.063)	(0.079)	(0.136)
Villagization × Birth Year 1960-1963	0.135	0.170	0.161	0.126	0.099	0.104	0.161	0.053
	(0.123)	(0.149)	(0.148)	(0.128)	(0.145)	(0.103)	(0.141)	(0.202)
Villagization × Birth Year 1964-1967	0.120	0.223*	0.217**	0.195*	0.167	0.196*	0.231*	0.082
	(0.098)	(0.122)	(0.104)	(0.106)	(0.101)	(0.105)	(0.124)	(0.138)
Villagization × Birth Year 1968-1971	0.176	0.284**	0.275**	0.255***	0.217**	0.258***	0.287**	0.190
	(0.109)	(0.107)	(0.103)	(0.094)	(0.105)	(0.084)	(0.110)	(0.139)
Villagization × Birth Year 1972-1975	0.062	0.139	0.132	0.108	0.083	0.110*	0.136	0.049
	(0.090)	(0.097)	(0.110)	(0.089)	(0.098)	(0.062)	(0.093)	(0.149)
Villagization × Birth Year 1976-1979	0.095	0.159	0.147	0.128	0.086	0.143	0.168*	0.074
	(0.090)	(0.108)	(0.122)	(0.119)	(0.124)	(0.112)	(0.098)	(0.191)
Villagization × Birth Year 1980-1983	0.016	0.060	0.051	0.027	0.029	0.026	0.057	-0.073
	(0.081)	(0.103)	(0.106)	(0.085)	(0.093)	(0.078)	(0.102)	(0.148)
Villagization × Birth Year 1984-1987	0.020	0.032	0.027	0.011	-0.029	0.003	0.009	-0.091
	(0.099)	(0.130)	(0.134)	(0.125)	(0.134)	(0.105)	(0.107)	(0.160)
Observations	1,797	1,797	1,797	1,797	1,797	1,797	1,797	1,797
Number of clusters	52	52	52	52	52	52	52	52
R-squared	0.087	0.111	0.115	0.118	0.115	0.141	0.122	0.125
District FE	✓	✓	✓	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓
Zone-Cohort FE		✓	✓	✓	✓	✓	✓	✓

Notes: The unit of observation is an individual Afrobarometer respondent i in district d belonging to birth cohort t. All columns regress the measure of an individual's current national identity on the interactions between that individual's district-level measure of historical villagization and dummies that indicate the individual's cohort, controlling for the 1967 district primary school enrollment rate interacted with the cohort dummies, survey year fixed effects, and the other controls and fixed effects indicated in the table. The un-interacted variables are included in the fixed effects in all columns. The dependent variable is from Afrobarometer rounds 3 and 4 (2005-2008), and was originally on a five-level Likert scale; it has been recoded from 0 to 1 in 1/4 increments (1 = respondent identifies only with the nation as a whole; 0 = respondent identifies only with their ethnic group). The reference group is the cohort born in 1948-1951. The villagization variable is the share of the respondent's current district's rural population that lived in registered government villages by 1978 according to the 1978 Tanzania Population Census. All controls, interacted with the individual-level cohort fixed effects, are at the district level and based on the 1967 Census (except weather shocks and geographical features). Geographical features include latitude, longitude, altitude and slope. Weather shocks are measured as z-scores of rainfall during the planting season in 1974, 1975 and 1976 (main period of mandatory villagization), censored at 0 to capture droughts. The weather data is based on readings from 108 weather stations which are spatially interpolated at the district level using Kriging. The villagization and outcome variables are standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. All regressions are weighted using the survey weights provided by Afrobarometer. Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by * p < 0.10, ** p < 0.05, *** p < 0.01.

Table A.6: The Effect of *Ujamaa* on National Identity, Robustness

	Dependent Variable: National vs. Ethnic Identity											
	baseline	unweighted	control for respondent beliefs about surveyor	rural sample only	dropping districts with highest migration rates	migration bounding exercise conservative case	migration bounding exercise alternative case	pca control	pca control	ordered probit	ordered logit	too young vs too old
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Villagization × Treated Cohort	0.226***	0.179***	0.215***	0.326**	0.224**	0.144**	0.164**	0.226***	0.154**	0.520**	0.861**	
	(0.081)	(0.060)	(0.078)	(0.126)	(0.088)	(0.067)	(0.070)	(0.080)	(0.074)	(0.206)	(0.400)	
Villagization × 1976-1987 Cohort												0.071
												(0.065)
Observations	849	849	849	710	668	849	849	849	785	849	849	936
Number of clusters	52	52	52	47	38	52	52	52	48	52	52	52
R-squared	0.139	0.123	0.144	0.167	0.136	0.133	0.134	0.139	0.135			0.113
District FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Zone-Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: The unit of observation is an individual Afrobarometer respondent i in district d belonging to school cohort t. Columns (1) to (11) regress the measure of an individual's current national identity on the interaction between that individual's district-level measure of historical villagization and a dummy that indicates whether the respondent is in the treated cohort, controlling for the 1967 district primary school enrollment rate interacted with the cohort dummy, survey year fixed effects, and the fixed effects indicated in the table. The un-interacted variables are included in the fixed effects in all columns. Column (12) uses the same specification, except that the treated cohort dummy takes value 1 for the 1976-1987 birth cohort. The dependent variable is from Afrobarometer rounds 3 and 4 (2005-2008), and was originally on a five-level Likert scale; it has been recoded from 0 to 1 in 1/4 increments (1 = respondent identifies only with the nation as a whole; 0 = respondent identifies only with their ethnic group). Treated cohort is a dummy that equals 1 if the respondent was born in 1960-1971. In all columns, the reference group is the cohort born in 1948-1959. The villagization variable is the share of the respondent's current district's rural population that lived in registered government villages by 1978 according to the 1978 Tanzania Population Census. The villagization and outcome variables are standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation, except in columns (10) and (11) in which the outcome is left discrete, not standardized. All regressions are weighted using the survey weights provided by Afrobarometer, except column (2) which does not apply the survey weights. Column (3) controls for a dummy variable that indicates whether the respondent erroneously believes that the Afrobarometer surveyor was sent by a Tanzanian government entity. Column (4) includes only rural respondents. Column (5) drops all districts with either an in- or an out-migration rate in the highest deciles. Columns (6) and (7) are explained in Appendix B. Column (8) includes also the first principal component of all controls added in columns (3) to (8) of Table 2 and columns (4) to (10) of Table 3. Column (9) includes the first principal component obtained including also national identity in 1967. Column (10) uses an ordered probit regression. Column (11) uses an ordered logit regression. The displayed coefficient shows the differential change in the ordered log-odds of reporting a higher level of the outcome variable, given a one-standard deviation increase in the villagization measure, for the treated cohort relative to the control one. The interpretation of the coefficient in column (10) is similar, but in a different scale as it is expressed in units of the probit index. Column (12) compares the 1976-1987 cohort to the 1948-1959 one, using only observations from these cohorts. Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7: The Effect of *Ujamaa* on Primary Education and Literacy in Swahili

Dep. Variable:	Primary School (1)	Swahili Literate (2)
Villagization × Treated Cohort	0.009*	0.008*
	(0.005)	(0.005)
Observations	562,722	653,272
Number of clusters	52	52
R-squared	0.152	0.085
District FE	✓	✓
Cohort FE	✓	✓
Zone-Cohort FE	✓	✓

Notes: The unit of observation is an individual IPUMS 1988 respondent i in district d belonging to school cohort t. The data includes the districts in the main sample as in Table 2. The dependent variable in column (1) is a dummy for whether the respondent completed primary school and in column (2) a dummy for whether the respondent is literate in Swahili. All columns regress the dependent variable on the interaction between that individual's district-level measure of historical villagization and a dummy that indicates whether the individual is in the treated cohort, controlling for the 1967 district primary school enrollment rate interacted with the cohort dummy, and the fixed effects indicated in the table. The un-interacted variables are included in the fixed effects in all columns. Respondents with partial primary schooling are excluded from the sample in column (1) and those who are still attending school are excluded from the sample in both columns (1) and (2). Treated cohort is a dummy that equals 1 if the respondent was born in 1960-1971. The reference group is the cohort born in 1948-1959. The villagization variable is the share of the respondent's current district's rural population that lived in registered government villages by 1978 according to the 1978 Tanzania Population Census. The villagization variable is standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. All regressions are weighted using the survey weights provided by IPUMS. Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.8: The Effect of *Ujamaa* on Occupation

	Dependent Variable: Occupation			
	government and parastatal (1)	private sector (2)	agriculture (3)	self-employed (4)
Villagization × Treated Cohort	-0.004	0.005	0.024	-0.016
	(0.005)	(0.006)	(0.021)	(0.020)
Observations	1,599	1,599	1,599	1,599
Number of clusters	52	52	52	52
R-squared	0.053	0.050	0.142	0.111
District FE	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓
Zone-Cohort FE	✓	✓	✓	✓

Notes: The unit of observation is an individual Tanzania National Panel Survey Round 1 (2008) respondent i in district d belonging to school cohort t. The data includes the districts in the main sample as in Table 2. All columns regress a dummy for the occupation listed in the column head on the interaction between an individual's district-level measure of historical villagization and a dummy that indicates whether the individual is in the treated cohort, controlling for the 1967 district primary school enrollment rate interacted with the cohort dummy, survey year fixed effects, and the fixed effects indicated in the table. The un-interacted variables are included in the fixed effects in all columns. Treated cohort is a dummy that equals 1 if the respondent was born in 1960-1971. The reference group is the cohort born in 1948-1959. We exclude a small number of respondents for which the birth date and age do not match in the original data. The villagization variable is the share of the respondent's current district's rural population that lived in registered government villages by 1978 according to the 1978 Tanzania Population Census. The villagization variable is standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. All regressions are weighted using the household survey weights provided by the data. Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.9: The Effect of *Ujamaa* on Economic Outcomes

Dep. Variable:	HH Consumption	HH Expenditures	Cash Income Job	Wealth Index	National Identity	
	(1)	(2)	(3)	(4)	(5)	(6)
Villagization × Treated Cohort	-0.184***	-0.034	-0.120**	-0.099		
	(0.057)	(0.033)	(0.050)	(0.072)		
Cash Income Job					0.032	
					(0.037)	
Wealth Index						0.073*
						(0.041)
Observations	958	958	868	859	849	841
Number of clusters	52	52	52	52	52	52
R-squared	0.137	0.095	0.187	0.178	0.134	0.138
District FE	✓	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓	✓
Zone-Cohort FE	✓	✓	✓	✓	✓	✓
Source	TNPS	TNPS	Afrobarometer	Afrobarometer	Afrobarometer	Afrobarometer

Notes: The unit of observation is an individual Tanzania National Panel Survey Round 1 (2008) respondent, in columns (1) and (2), and an Afrobarometer respondent, in columns (3) to (6), i in district d belonging to school cohort t. The data includes the districts in the main sample as in Table 2. The sample in columns (1) and (2) is restricted to household heads. Columns (1) to (4) regress the outcome stated in the column heads on the interaction between an individual's district-level measure of historical villagization and a dummy that indicates whether that individual is in the treated cohort, controlling for the 1967 district primary school enrollment rate interacted with the cohort dummy, survey year fixed effects, and the fixed effects indicated in the table. The un-interacted variables are included in the fixed effects in all columns. Column (5) regresses the national identity measure on having a job that pays cash income and column (6) on a wealth index, controlling for the 1967 district primary school enrollment rate interacted with the cohort dummy, survey year fixed effects, and the fixed effects indicated in the table. Household consumption is total annual real household consumption per adult equivalent. Household expenditures are total annual nominal household expenditures on furnishings and household items. Cash Income Job is from Afrobarometer rounds 3 and 4 (2005-2008), and was originally on a three-level scale; it has been recoded from 0 to 1 in 1/2 increments (1 = respondent has a full time job that pays cash income; 1/2 = respondent has a part time job that pays cash income; 0 = respondent does not have a job that pays cash income). The wealth index is from Afrobarometer rounds 3 and 4 (2005-2008), and is obtained from the first principal component of the questions asking about the frequency of the household not having enough food, clean water, medicines, cooking fuel, cash income, and about ownership of a radio, a tv and a motorbike. These questions are present in both rounds 3 and 4. Treated cohort is a dummy that equals 1 if the respondent was born in 1960-1971. The reference group is the cohort born in 1948-1959. We exclude a small number of respondents for which the birth date and age do not match in the original data. The villagization variable is the share of the respondent's current district's rural population that lived in registered government villages by 1978 according to the 1978 Tanzania Population Census. The national identity outcome is from Afrobarometer rounds 3 and 4 (2005-2008), and was originally on a five-level Likert scale; it has been recoded from 0 to 1 in 1/4 increments (1 = respondent identifies only with the nation as a whole; 0 = respondent identifies only with their ethnic group). The villagization, national identity, cash income job and wealth index measures are standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. All regressions are weighted using the survey weights provided by TNPS, for columns (1) and (2), and Afrobarometer, for columns (3) to (6). Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.10: The Effects of *Ujamaa* on State Legitimacy and Accountability

Dep. Variable:	respect authority (1)	support one-party rule (2)	see gov. as parent (3)	trust gov. newspaper (4)	trust gov. TV/radio (5)	took action to hold gov. accountable (6)	trust own ethnic group (7)	trust other ethnic group (8)
Villagization × Treated Cohort	0.169**	0.097	0.230**	0.097	0.205***	-0.051	0.055	-0.141
	(0.071)	(0.064)	(0.111)	(0.102)	(0.069)	(0.031)	(0.097)	(0.096)
Observations	853	846	353	425	434	861	493	492
Number of clusters	52	52	47	45	45	52	45	45
R-squared	0.085	0.144	0.174	0.184	0.165	0.153	0.171	0.152
District FE	✓	✓	✓	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓
Zone-Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓

Notes: The unit of observation is an individual Afrobarometer respondent i in district d belonging to school cohort t. All columns regress the outcome stated in the column heads on the interaction between an individual's district-level measure of historical villagization and a dummy that indicates whether that individual is in the treated cohort, controlling for the 1967 district primary school enrollment rate interacted with the cohort dummy, survey year fixed effects, and the fixed effects indicated in the table. The un-interacted variables are included in the fixed effects in all columns. The dependent variables are from Afrobarometer rounds 3 and/or 4 (2005-2008) and are as follows (recoded from Likert scale). (1) "Choose Statement 1 or Statement 2. Do you agree or agree very strongly? Statement 1: Citizens should be more active in questioning the actions of leaders. Statement 2: In our country, citizens should show more respect for authority." The dependent variable registers agreement with Statement 2.; (2) "There are many ways to govern a country. Would you disapprove or approve of the following alternatives? Only one political party is allowed to stand for election and hold office." The dependent variable registers approval with this option.; (3) "Choose Statement 1 or Statement 2. Do you agree or agree very strongly? Statement 1: People are like children; the government should take care of them like a parent. Statement 2: Government is like an employee; the people should be the bosses who control the government." The dependent variable registers agreement with Statement 1.; (4) "How much do you trust each of the following, or haven't you heard enough about them to say: Government (relative to independent) newspapers?"; (5) "Government (relative to independent) broadcasting service (TV / radio)?""; (6) "Here is a list of actions that people sometimes take as citizens. For each of these, please tell me whether you, personally, have done any of these things during the past year" Got together with others to raise an issue"; (7) "How much do you trust each of the following types of people: "People from your own ethnic group?". (8) "People from other ethnic groups?". Questions (4), (5), (7) and (8) were asked only in Round 3, question (3) only in Round 4. Treated cohort is a dummy that equals 1 if the respondent was born in 1960-1971. The reference group is the cohort born in 1948-1959. The villagization variable is the share of the respondent's current district's rural population that lived in registered government villages by 1978 according to the 1978 Tanzania Population Census. The villagization and outcome variables, except the outcome of column (6) which is binary, are standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. All regressions are weighted using the survey weights provided by Afrobarometer. Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.11: The Effect of *Ujamaa* on Public Goods, Gender Norms, Voting

Dep. Variable:	Supports Free Schooling (1)	Agrees Women Equal Rights (2)	Agrees Women Equal Election Chances (3)	Supports CCM (4)	CCM Support in Elections (5)
Villagization × Treated Cohort	0.236***	-0.091	0.038	0.013	
	(0.049)	(0.152)	(0.104)	(0.025)	
Villagization					0.190**
					(0.077)
Observations	490	493	493	699	112
Number of clusters	45	45	45	52	56
R-squared	0.163	0.115	0.156	0.154	0.212
District FE	✓	✓	✓	✓	
Cohort FE	✓	✓	✓	✓	
Zone FE	✓	✓	✓	✓	✓
Zone-Cohort FE	✓	✓	✓	✓	
Source	Afrobarometer	Afrobarometer	Afrobarometer	Afrobarometer	Electoral data

Notes: In columns (1) to (4), the unit of observation is an individual Afrobarometer respondent i in district d belonging to school cohort t. All columns regress the outcome stated in the column heads on the interaction between an individual's district-level measure of historical villagization and a dummy that indicates whether that individual is in the treated cohort, controlling for the 1967 district primary school enrollment rate interacted with the cohort dummy, survey year fixed effects, and the fixed effects indicated in the table. The un-interacted variables are included in the fixed effects in all columns. These regressions are weighted using the survey weights provided by Afrobarometer. The dependent variables in columns (1) to (3) are from Afrobarometer Round 3 (2005). They are defined as follows (recoded from Likert scale). (1) "Which of the following statements is closest to your view? Choose Statement A or Statement B. A: It is better to have free schooling for our children, even if the quality of education is low. B: It is better to raise educational standards, even if we have to pay school fees." The dependent variable registers agreement with Statement A.; (2) "Which of the following statements is closest to your view? Choose Statement A or Statement B. A: In our country, women should have equal rights and receive the same treatment as men do. B: Women have always been subject to traditional laws and customs, and should remain so." The dependent variable registers agreement with Statement A.; (3) "Which of the following statements is closest to your view? Choose Statement A or Statement B. A: Women should have the same chance of being elected to political office as men. B: Men make better political leaders than women, and should be elected rather than women." The dependent variable registers agreement with Statement A. The dependent variable in column (4) is from Afrobarometer rounds 3 (2005) and 4 (2008) and is coded as a dummy for whether the respondent feels close to a particular party and this party is Chama cha Mapinduzi (CCM). In column (5), the unit of observation is a district d in year t. This column regresses support for the CCM on the district-level measure of historical villagization, controlling for the 1967 district primary school enrollment rate, election year fixed effects, and the fixed effects indicated in the table. CCM support is defined as the product of turnout and CCM presidential vote share from the 2000 and 2005 presidential elections in Tanzania (data from Carlitz (2017)), aggregated at the level of 1967 districts, averaging over constituencies. Treated cohort is a dummy that equals 1 if the respondent was born in 1960-1971. The reference group is the cohort born in 1948-1959. The villagization variable is the share of the respondent's current district's rural population that lived in registered government villages by 1978 according to the 1978 Tanzania Population Census. The villagization measure and outcomes in columns (1), (2), (3) and (5) are standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

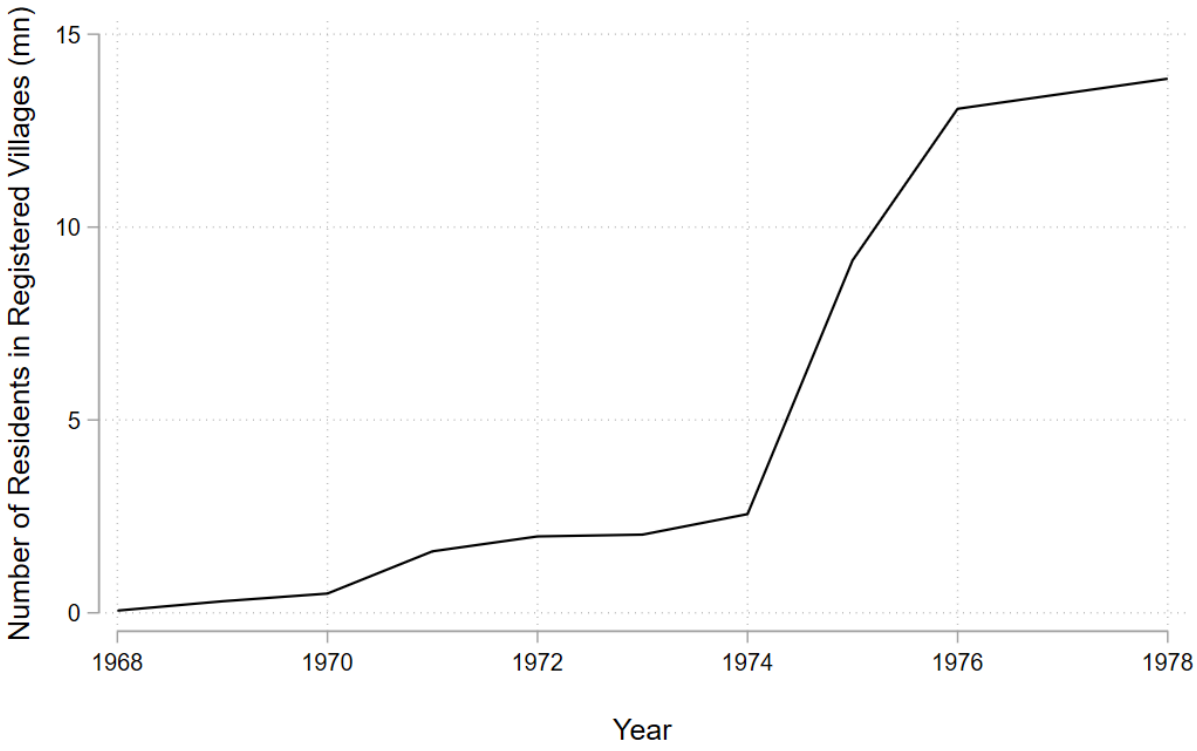
Table A.12: Correlations Between Villagization and District Characteristics in 1967

1967 district characteristic	correlation	p-value
Primary school enrollment rate in 1967 (per '000 inhabitants)	0.27	0.07
Distance to Dar es Salaam in kilometers	-0.08	0.69
District revenue per capita in 1967 ('000 shs)	-0.03	0.78
Ethnolinguistic fractionalization in 1967	-0.09	0.34
Centroid latitude	-0.08	0.81
Centroid longitude	0.14	0.46
Average altitude in meters	-0.17	0.42
Average slope	-0.15	0.13
Hospital beds in 1967 (per '000.000 inhabitants)	0.04	0.81
Drought in 1974 (censored z-score)	-0.06	0.74
Drought in 1975 (censored z-score)	0.05	0.74
Drought in 1976 (censored z-score)	0.44	0.06
National Identity in 1967	-0.07	0.61
Share urban in 1967 (% of population)	0.03	0.84
Population density in 1967	0.22	0.27
GDP per capita in 1967 (mln shs)	0.13	0.48
Tot. market agricultural production per capita in 1967 ('000 shs)	0.23	0.29
Distance to Uganda (km)	0.51	0.32
Number of missions per capita	-0.31	0.26
Total length of roads over district area (km/km2)	0.16	0.4

Notes: This table shows pairwise correlations between villagization (share of total rural district population living in registered villages in 1978) and the variables indicated in the table, conditional on zone fixed effects. The reported p-values are robust to hetero-skedasticity. Units of observations are 1967 districts and the data includes the districts in the main sample as in Table 2. The variables and their sources are described in detail in Section 3 and Appendix A.

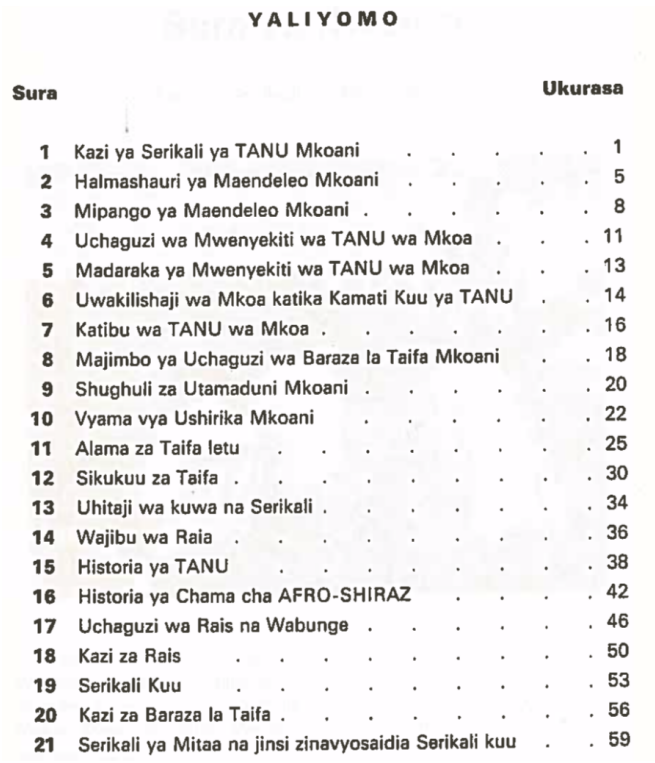
Appendix G. Additional Figures

Figure A.1: Population in Registered Villages Over Time



Notes: Authors' illustration is based on data from Shao (1982), which has been digitized by Delesalle (2021). The data for the year 1977 was missing from the original source. To estimate the number of residents in registered villages for that year, we applied a linear interpolation between the values for 1976 and 1978. The implementation of the villagization program began after the issuance of Presidential Circular No. 1 in 1969, accelerated in 1974 when villagization was declared mandatory, and formally ended with the repeal of the Villages and *Ujamaa* Villages Act in 1982.

Figure A.2: Political Education Textbook for 4^{th} Graders



YALIYOMO

Sura		Ukurasa
1	Kazi ya Serikali ya TANU Mkoani	1
2	Halmashauri ya Maendeleo Mkoani	5
3	Mipango ya Maendeleo Mkoani	8
4	Uchaguzi wa Mwenyekiti wa TANU wa Mkoa	11
5	Madaraka ya Mwenyekiti wa TANU wa Mkoa	13
6	Uwakilishaji wa Mkoa katika Kamati Kuu ya TANU	14
7	Katibu wa TANU wa Mkoa	16
8	Majimbo ya Uchaguzi wa Baraza la Taifa Mkoani	18
9	Shughuli za Utamaduni Mkoani	20
10	Vyama vya Ushirika Mkoani	22
11	Alama za Taifa letu	25
12	Sikukuu za Taifa	30
13	Uhitaji wa kuwa na Serikali	34
14	Wajibu wa Raia	36
15	Historia ya TANU	38
16	Historia ya Chama cha AFRO-SHIRAZ	42
17	Uchaguzi wa Rais na Wabunge	46
18	Kazi za Rais	50
19	Serikali Kuu	53
20	Kazi za Baraza la Taifa	56
21	Serikali ya Mitaa na jinsi zinavyosaidia Serikali kuu	59

CONTENTS

1. The work of the TANU government in the regions
2. Regional development councils
3. Regional development plans
4. Election of TANU leaders in regions
5. Powers of TANU leaders in regions
6. Regional representatives in TANU Central Committee
7. TANU's provincial secretary
8. Constituencies for national elections
9. Services in the regions
10. Participatory committees in the regions
11. Symbols of our nation
12. National Holidays
13. The need to support the government
14. Civic duties
15. History of TANU
16. History of the AFRO-SHIRAZ party
17. Election of the President and Parliament
18. Work of the President
19. Central government
20. Work of the central council
21. Local governments and how they help the central government

Notes: This image depicts the contents page of a political education textbook for 4^{th} graders. The figure on the left shows the contents page in Swahili, while the figure on the right provides the equivalent English translation by the authors. *Sura* translates to "chapter" and *Ukurasa* to "pages". Items that are underlined highlight specific elements of nation-building. *Source*: Textbook titled *Elimu ya Siasa*, translated as *Political Education* (United Republic of Tanzania National Department of Education, 1971a).

Figure A.3: Excerpts from Swahili Textbook

1. UMOJA NI NGUVU

Mtu anavunja kitita cha vijiti 10

(a) *"Unity is strength"*

18. UHURU NA JAMHURI

Mwakapesa tangu alipoanza shule, aliwaona watu wakisherehekea sikukuu ya "Uhuru na Jamhuri". Sherehe zilifanywa pia na watoto wa shule. Watu waliimba nyimbo zilizosifu **ukombozi** wa nchi yetu.

Mwakapesa hakuchaguliwa kuingia katika mashindano ya sikukuu mpaka alipoingia darasa la tano. Mwalimu wake alitunga nyimbo za kulaani utawala wa kigeni na za kuwasifu viongozi wa Tanzania. Mwakapesa alichaguliwa kuongoza kundi la waimbaji. Alikuwa hodari sana wa kuimbisha wenzake na wala hakuogopa kusimama mbele za watu.

Desemba tisa ilipofika, shule ya Mpuguso ilikwenda kwenye mashindano. Uwanja wa mashindano ulikuwa mbele ya Jumba la Maendeleo. Kabla ya kuimba, waimbaji walijitayarisha nyuma ya jengo ili wapate nafasi kubwa ya kutembea kigwaride.

(b) *"Independence day"*

Notes: These images depict pages from Chapter 1 (left), titled *Unity is Strength*, and Chapter 18 (right), titled *Independence and the Republic*, of the Swahili textbook *Tujifunze Lugha Yetu: Kitabu Cha Saba* (translated as *Let us learn our language: Class Seven*) (United Republic of Tanzania National Department of Education, 1971b).

Figure A.4: Population Census 1978, Share of Rural Population in Official Registered Villages

TABLE 1 POPULATION BY AGE AND SEX: VILLAGES/WARDS, DISTRICTS AND REGIONS, 1978 PAGE 45

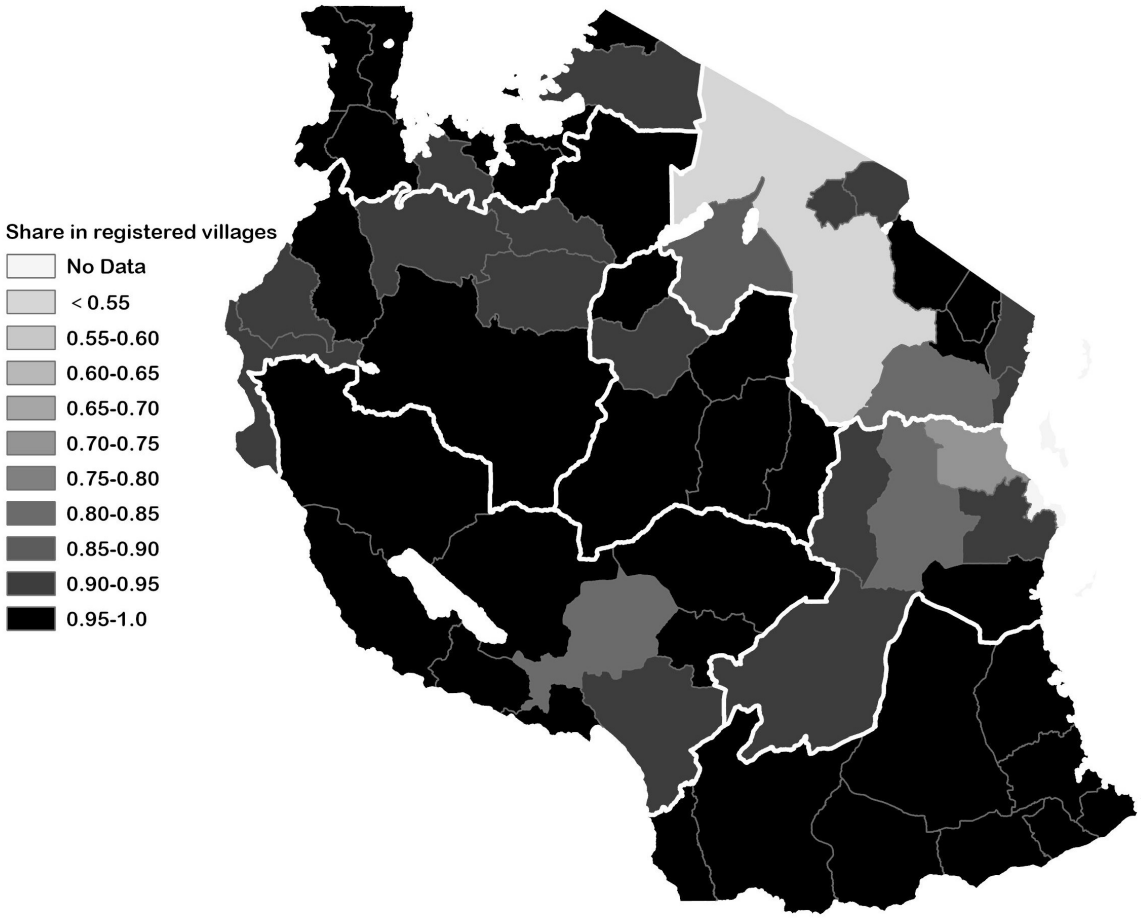
REGION: 02 ARUSHA
DISTRICT: 1 MONDULI

AREA/POP. CATEGORY	SEX	0-4	5-9	10-14	15-24	25-34	35-44	45-54	55-64	65+	TOTAL
REGISTERED VILLAGES	MALES	5676	4925	3760	4881	3649	2803	1905	1275	1225	30095
	FEMALES	6163	4749	3221	5547	4489	2427	1581	992	920	30089
	TOTAL	11839	9674	6981	10428	8138	5230	3486	2265	2143	60184
NONREGISTERED VILLAGES	MALES	2530	2368	1355	1938	1959	1409	993	705	542	13799
	FEMALES	2820	2237	1010	2779	2204	1332	819	534	581	14316
	TOTAL	5350	4605	2365	4717	4163	2741	1812	1239	1123	28115
MIGRATORY POPULATION	MALES	2192	2009	1447	1543	1520	1176	678	443	512	11520
	FEMALES	2578	1873	876	2323	2009	1084	653	444	396	12236
	TOTAL	4770	3882	2323	3866	3529	2260	1331	887	908	23756
INSTITUTION. POP. [1]	MALES	241	117	95	1159	982	521	184	57	15	3371
	FEMALES	253	127	103	328	194	67	36	15	1	1122
	TOTAL	494	244	198	1487	1176	588	220	70	16	4493
TOTAL RURAL AREAS	MALES	10639	9419	6657	9521	8110	5909	3760	2478	2292	58785
	FEMALES	11814	8986	5210	10977	8896	4910	3089	1983	1898	57763
	TOTAL	22453	18405	11867	20498	17006	10819	6849	4461	4190	116548
TOTAL URBAN AREAS	MALES	200	132	85	427	276	112	39	21	11	1303
	FEMALES	193	124	110	273	142	51	30	12	10	945
	TOTAL	393	256	195	700	418	163	69	33	21	2248
TOTAL DISTRICT	MALES	10839	9551	6742	9948	8386	6021	3799	2499	2303	60088
	FEMALES	12007	9110	5320	11250	9038	4961	3119	1995	1908	58708
	TOTAL	22846	18661	12062	21198	17424	10982	6918	4494	4211	118796

[1] IF A MIXED WARD HAD AN URBAN COMPONENT OF INSTITUTIONAL POPULATION, THIS COMPONENT WAS INCLUDED HERE.

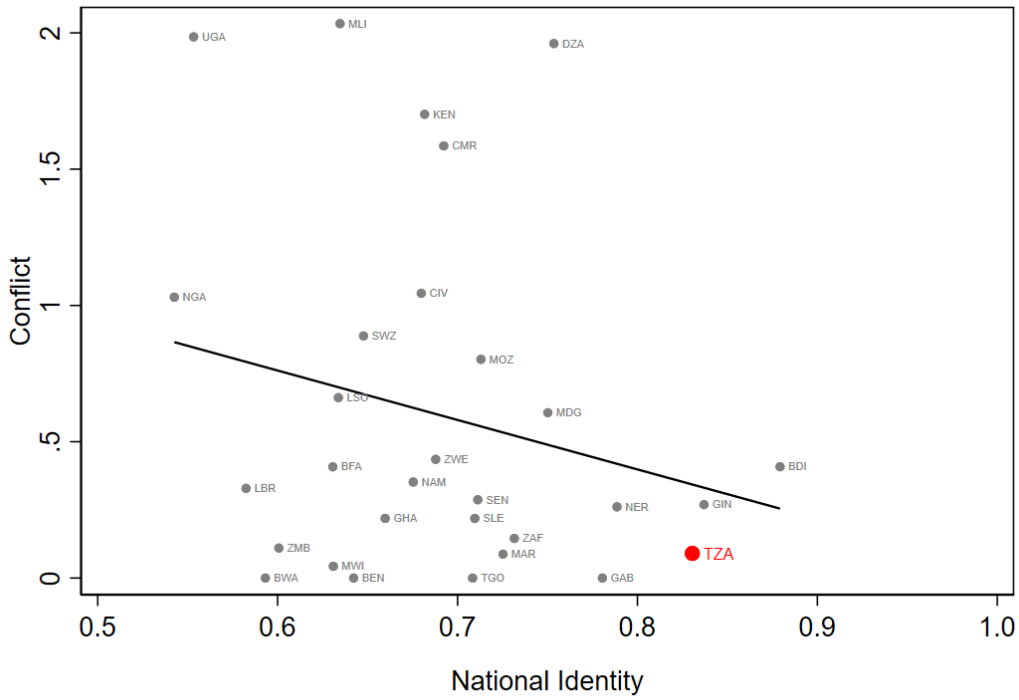
Notes: This image shows an excerpt from the 1978 Tanzania Population Census (Bureau of Statistics, Ministry of Planning and Economic Affairs, 1981), retrieved from the Herskovits Library of African Studies at Northwestern University.

Figure A.5: Distribution of Villagization



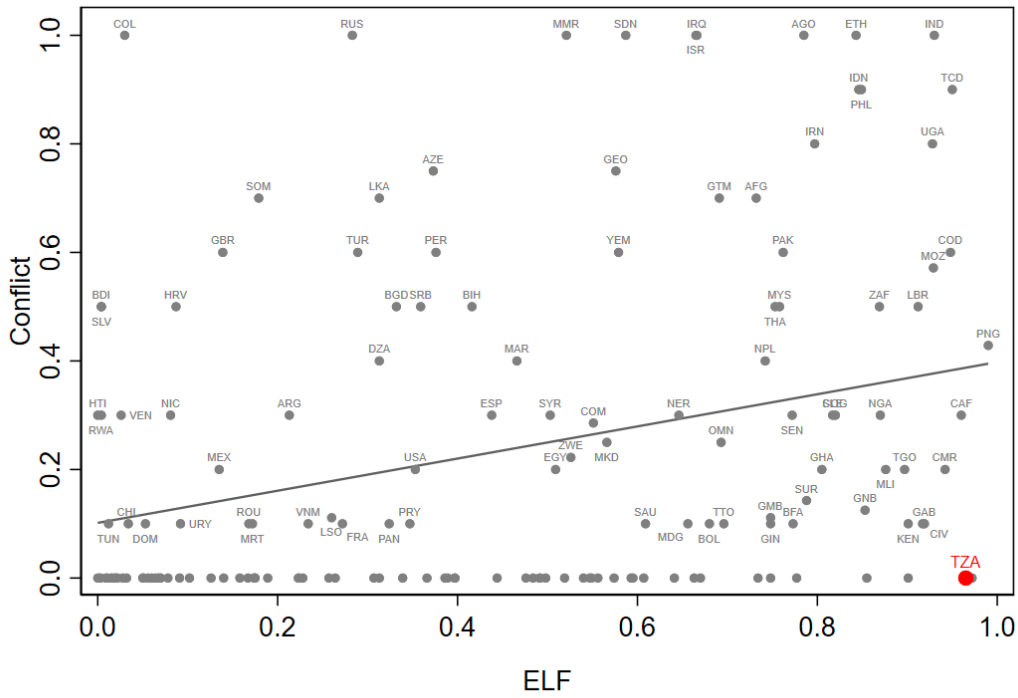
Notes: This figure depicts our district-level measure of villagization, defined as the share of the rural population living in officially registered villages according to the 1978 Tanzania Population Census (at 1967 district borders). The population not living in official registered villages includes those in unregistered villages and individuals living outside villages (scattered and migratory population). Thick white lines indicate geographic zone boundaries. Dar es Salaam (Mzizima) and islands (Zanzibar, Pemba, Mafia) are excluded from the sample. Sources: Authors' illustration. Base map from The Humanitarian Data Exchange (HDX) by United Nations OCHA ROSA (https://data.humdata.org/, last accessed on May 26, 2024).

Figure A.6: National Identity and Conflict in Africa



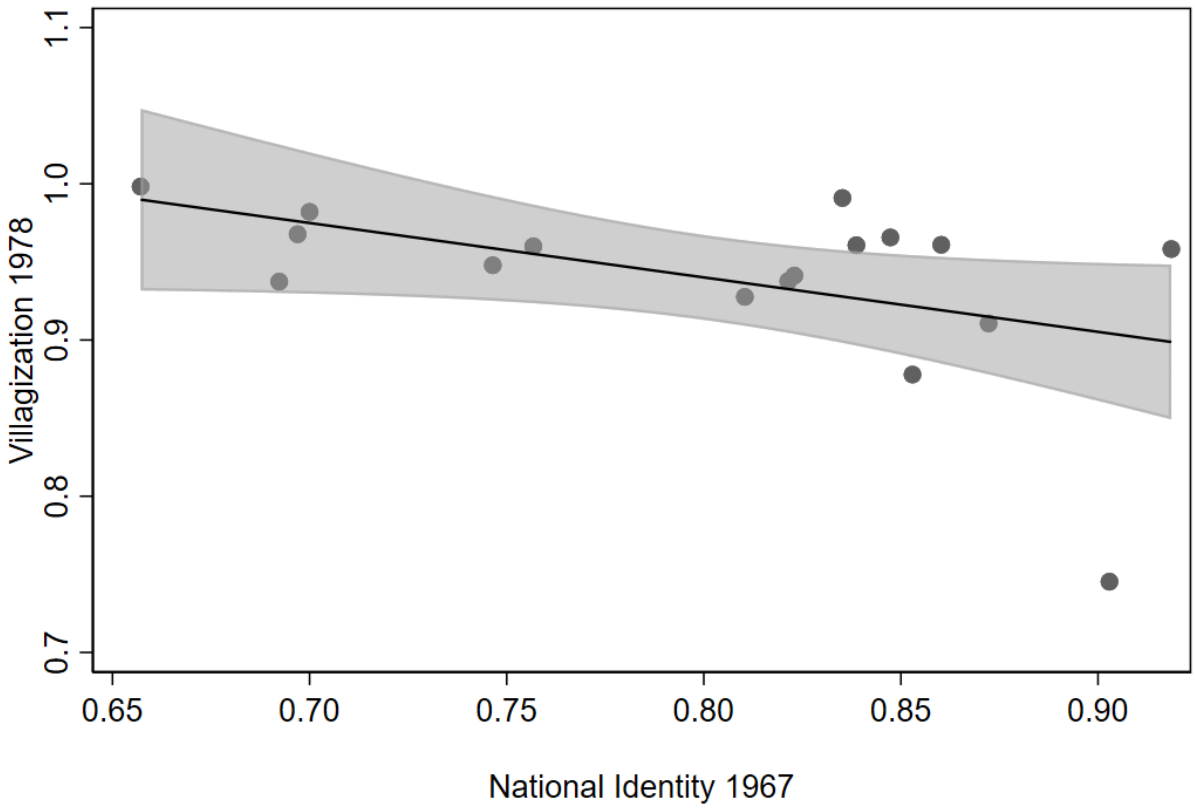
Notes: This plot illustrates the correlation between the mean number of battles per 1 million inhabitants (defined as "conflict") and the average expressed feelings of national identity (as opposed to ethnic identity) for 33 African countries between 2005 and 2014. Conflict data is from ACLED (Raleigh et al., 2023), and population data is from the World Development Indicators (World Bank, 2020). Country averages of national identity are taken from Afrobarometer Rounds 3-6.

Figure A.7: Conflict and Ethnolinguistic Fractionalization



Notes: This plot illustrates the correlation between conflict (defined as the incidence of internal armed conflict with at least 25 deaths in the last 5 years based on data from Gleditsch et al. (2002)) and ethnolinguistic fractionalization (index based on data from Fearon (2003)) at the country level between 1960 and 2008. The data are sourced from Esteban et al. (2012), which provides further details.

Figure A.8: National Identity in 1967 and Villagization in 1978



Notes: This plot shows the correlation (with 95% confidence interval) between region-level villagization in 1978 and average feelings of national identity in 1967. Villagization is defined as the fraction of the rural region population living in official registered villages, according to the 1978 Tanzania Population Census. Region-level villagization is computed as the within-region average of district-level villagization in 1967. The data on national identity is based on a nationwide survey of almost 3,000 Tanzanian secondary school students in 1967 by Prewitt et al. (1970). The original variable is coded as 1 if a respondent feels that the nation is more important than the tribe, and 0 otherwise.

Online Appendix References

Barker, J. S. (1974). "Ujamaa" in Cash-Crop Areas of Tanzania: Some Problems and Reflections. *Journal of African Studies 1*(4), 441.

Benshaul-Tolonen, A., S. Aguilar-Gomez, N. Heller Batzer, R. Cai, and E. C. Nyanza (2020). Period teasing, stigma and knowledge: A survey of adolescent boys and girls in Northern Tanzania. *PLoS One 15*(10), e0239914.

Boesen, J., B. S. Madsen, and T. Moody (1977). *Ujamaa – Socialism from above*. Scandinavian Institute of African Studies, Uppsala.

Bureau of Statistics, Ministry of Planning and Economic Affairs (1981). *1978 Population Census*. Dar es Salaam: United Republic of Tanzania.

Carlitz, R. D. (2017). Money flows, water trickles: Understanding patterns of decentralized water provision in Tanzania. *World Development 93*, 16–30.

Chamie, M. (1983). *National, institutional, and household factors affecting young girls' school attendance in developing societies*. International Center for Research on Women.

Delesalle, E. (2021). The effect of the universal primary education program on consumption and on the employment sector: Evidence from tanzania. *World Development 142*, 105345.

Dell, M., B. F. Jones, and B. A. Olken (2014). What do we learn from the weather? The new climate-economy literature. *Journal of Economic Literature 52*(3), 740–98.

Ergas, Z. (1980). Why did the Ujamaa village policy fail? Towards a global analysis. *The Journal of Modern African Studies 18*(3), 387–410.

Esteban, J., L. Mayoral, and D. Ray (2012). Ethnicity and conflict: An empirical study. *American Economic Review 102*(4), 1310–42.

Fearon, J. D. (2003). Ethnic and cultural diversity by country. *Journal of economic growth 8*, 195–222.

Gleditsch, N. P., P. Wallensteen, M. Eriksson, M. Sollenberg, and H. Strand (2002). Armed conflict 1946-2001: A new dataset. *Journal of peace research 39*(5), 615–637.

Green, D. P., D. W. Groves, C. Manda, B. Montano, and B. Rahmani (2023). A Radio Drama's Effects on Attitudes Toward Early and Forced Marriage: Results From a Field Experiment in Rural Tanzania. *Comparative Political Studies 56*(8), 1115–1155.

Hayuma, A. (1980). Dodoma: The planning and building of the new capital city of Tanzania. *Habitat International 5*(5-6), 653–680.

Hedde-von Westernhagen, C. and B. Becker (2022). Mapping Missions: New Data for the Study of African History: Social and Economic History. *Research Data Journal for the Humanities and Social Sciences 7*(1), 1–33.

Herbst, J. (2014). *States and Power in Africa: Comparative Lessons in Authority and Control*. Princeton University Press.

Hydén, G. (1980). *Beyond Ujamaa in Tanzania: Underdevelopment and an Uncaptured Peasantry*. University of California Press.

Jedwab, R. and A. Storeygard (2022, 06). The Average and Heterogeneous Effects of Transportation Investments: Evidence from Sub-Saharan Africa 1960–2010. *Journal of*

the European Economic Association 20(1), 1–38.

Jensen (1968). *Government of Tanzania District Statistics*. Dar es Salaam: Government Printers.

McHenry, D. E. (1979). *Tanzania's Ujamaa Villages. The Implementation of a Rural Development Strategy*. Institute of International Studies, University of California, Berkeley.

Nunn, N. and D. Puga (2012). Ruggedness: The blessing of bad geography in Africa. *Review of Economics and Statistics 94*(1), 20–36.

Osafo-Kwaako, P. (2012). *Essays in Economic History and Development*. Ph. D. thesis, Harvard University.

Prewitt, K., G. Von Der Muhll, and D. Court (1970). School Experiences and Political Socialization: A Study of Tanzanian Secondary School Students. *Comparative Political Studies 3*(2), 203–225.

Raleigh, C., R. Kishi, and A. Linke (2023). Political instability patterns are obscured by conflict dataset scope conditions, sources, and coding choices. *Humanities and Social Sciences Communications 10*(74).

Roberts, G. (2014). The Uganda–Tanzania War, the fall of Idi Amin, and the failure of African diplomacy, 1978–1979. *Journal of Eastern African Studies 8*(4), 692–709.

Scott, J. C. (1998). *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Yale University Press.

Shao, I. F. (1982). A neo-colony and its problems during the process of attempting to bring about socialist rural transformation: the case of Tanzania. In *Taamuli: A Political Science Forum*, Volume 12, pp. 29–46.

Sitari, T. (1983). Settlement changes in the Bagamoyo district of Tanzania as a consequence of villagization. *Fennia-International Journal of Geography 161*(1), 1–90.

Stabler, E. (1979). Kenya and tanzania: Strategies and realities in education and development. *African Affairs 78*(310), 33–56.

United Republic of Tanzania National Department of Education (1971a). *Elimu ya Siasa 6: Taifa na Serikali Wetu*. Tanzania Publishing House.

United Republic of Tanzania National Department of Education (1971b). *Tujifunze lugha yetu : kitabu cha saba*. Tanzania Publishing House.

World Bank (2020). *World Development Indicators 2020*. The World Bank.