# Auxiliary Analysis

# Estimates from ordered logit.

The baseline estimates were obtained using OLS. Nonetheless, the Afrobarometer respondents’ answers to the identification question are ordinal. In particular, the relevant answers on the variable are ordinal that take the values 1 for “ I feel only (R’s ethnic group)”, 2 for “I feel more (R’s ethnic group) than national ID”, 3 for “I feel equally national ID and (R’s ethnic group)”, 4 for “I feel more national ID than (R’s ethnic group)”, and 5 for “I feel only national ID”. An alternative strategy is to maintain the ordinal nature of the answers and instead estimate an ordered logit model. Thus, I check for the robustness of the results in Table 1 using an ordered logit model.

Table AS1 reports the baseline estimates of the paper using an ordered logit model. The estimated coefficients reported in the top panel of the table are negative and statistically significant. The estimates show that if an individual’s ancestors were more heavily exposed to early statehood, then he or she is less likely to identify in ethnic terms. Marginal effects from the ordered logit are reported in the lower panel of Table AS1. Each row of the panel reports the marginal effect for each of the five possible responses to the ethnic versus national identification question. According to the estimates in column 3 of Table AS1, being a member of a centralized ethnic group is associated with being 5.4 % more likely to identify only in national terms, 14.6 % more likely to identify in national than in ethnic terms, 3.3 % less likely to identify in ethnic than in national terms, and 2.9 % less likely to identify only in ethnic terms.

Column 1 of Table AS2 excludes Tanzania from the sample to show that Tanzania’s historical nation-building success does not entirely drive the results (Bandyopadhyay and Green, 2013; Eifert, Miguel, and Posner, 2010; Lieberman and Singh, 2012; Miguel, 2003),

To make sure that Tanzania’s historical nation-building success does not entirely drive the results (Bandyopadhyay and Green, 2013; Eifert, Miguel, and Posner, 2010; Lieberman and Singh, 2012; Miguel, 2003), column 1 of Table AS2 excludes Tanzania from the sample.

Table AS1: Ordered Logit Estimates of Precolonial Centralization on Identification

|  |  |  |  |
| --- | --- | --- | --- |
|  | (1) | (2) | (3) |
| Centralized Dummy [Estimated coefficient] | -0.352\*\* | -0.358\*\*\* | -0.349\*\*\* |
| Category of Responses & Marginal Effects after logit | (0.151) | (0.091) | (0.089) |
| Feels only National ID (ENI=1) | 0.080\*\* | 0.059\*\*\* | 0.054\*\*\* |
|  | (0.035) | (0.016) | (0.015) |
| Feels national ID than ethnic group (ENI=2) | 0.007\*\* | 0.015\*\*\* | 0.146\*\*\* |
|  | (0.003) | (0.004) | (0.0037) |
| Feels ethnic group equally as national ID (ENI=3) | -0.046\*\* | -0.011 | -0.066 |
|  | (0.019) | (0.009) | (0.008) |
| Feels ethnic group than national ID (ENI=4) | -0.024\*\* | -.033\*\*\* | -0.033\*\*\* |
|  | (0.011) | (.008) | (0.0082) |
| Feels only ethnic group (ENI=5) | -0.017\*\* | -0.029\*\*\* | -0.029\*\*\* |
|  | (0.007) | (0.008) | (0.0076) |
| Observations | 57,758 | 55,101 | 51,965 |
| Pseudo R-squared | 0.008 | 0.034 | 0.037 |
| Pre-treatment controls | No | Yes | Yes |
| Post-treatment Controls | No | No | Yes |
| Country FE | No | Yes | Yes |
| Wave FE | No | Yes | Yes |

*Notes:* Standard errors are clustered at the ethnicity level. \*\*\* p*<*0.01, \*\* p*<*0.05.

# Social desirability bias

Social desirability bias refers to the tendency of respondents to give answers that they believe are socially acceptable, even if they do not reflect their true feelings or beliefs. In the data, the “national = ethnic” response option dominates within the dependent variable. This could, to some extent, reflect social desirability bias or satisficing as respondents may not want to be seen adopting a potentially controversial position. Therefore, our results may be biased due to such social desirability bias. To check for this possibility, I dropped the “national = ethnic” category, and re-estimated the main specification. The result is reported in column 2 of Table AS2. The result remains robust. I also dropped the “national = ethnic” category and created a dummy variable that equals 1 if 1 *<*= *ENI <*= 2 or 0 if 4 *<*= *ENI <*= 5. The results of this exercise are reported in column 5 of Table 1 and in column 3 of Table AS2, where both the dependent and explanatory variables are binary dummies. Overall, the result remains robust irrespective of whether I use the original 5-point scale or the dummy measure of ethnic identification (ENI). Taken together, the findings suggest that social desirability bias is unlikely to significantly impact our results. This conclusion rests on the consistency of the estimated effect of political centralization on ethnic identification across alternative measures.

Note that the IV estimate differs from the OLS estimate. In column 5 of Table 1, the OLS estimate shows that descendants of precolonially centralized societies are approximately 5.5 percent less likely to prioritize their ethnicity over their national identity. However, the 2SLS estimate in column 3 of Table AS2 demonstrates that members of centralized groups are 9.5 percent less likely to prioritize their ethnic group over the nation. This discrepancy between the two estimates provides further support for the validity of our instrumental variable (IV) approach.

An alternative measure of political centralization is also utilized in this study. Specifically, in column 4 of Table AS2, I report the 2SLS results using the original 5-point measure of political centralization. The robustness of the results is maintained under this alternative measure.

# Colonial and Contemporary Factors

Thus far, the results focused on how precolonial centralization plays a role in influencing the salience of ethnicity. However, an increasing body of research suggests that colonial policies and present-day political factors have played a significant role in shaping the significance of ethnicity in modern Africa. In this section, I aim to demonstrate that the findings remain robust even when accounting for these colonial and contemporary factors. To conduct this analysis, I utilize the replication data from Robinson (2014) and McNamee (2019), as reported by McNamee (2019). The results of this exercise are reported in Table AS3 [[1]](#footnote-1).

Table AS2: Robustness Checks

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) |
| Centralized Dummy | -0.325\*\*\* | -0.412\*\* | -0.095\*\* |  |
|  | (0.126) | (0.163) | (0.047) |  |
| Centralization (5 scale) |  |  |  | -0.195\*\*(0.078) |
| Observations | 51,965 | 30,663 | 30,663 | 51,965 |
| R-squared | 0.016 | 0.024 | 0.027 | 0.009 |

*Notes:* In all columns, all controls are included, and the estimates are from 2SLS. Standard errors are clustered at the ethnicity level. \*\*\* p*<*0.01, \*\* p*<*0.05.

# The Influence of Indirectness in Colonial Rule

My previous argument has emphasized the impact of precolonial centralization on the importance of ethnicity, transcending the influence of colonial rule and its policies. Previously, I illustrated this point using the example of Ethiopia. However, it is important to acknowledge that the Ethiopian case cannot be generalized to the rest of Africa due to the absence of colonial influence. In other parts of Africa, colonialism left legacies that had lasting impacts. The British colonial administration predominantly employed indirect rule, which involved the establishment of native authorities and the maintenance of customary control over land. On the other hand, the French administration commonly implemented the direct rule, relying less on pre-existing traditional institutions and striving to promote a unifying republican ideology, aiming to integrate native Africans as potential French citizens. Recent research by McNamee (2019) provides further insights by highlighting the role of the degree of indirectness in colonial rule in shaping the salience of ethnicity. Therefore, it is crucial for the analysis to consider the level of indirectness in colonial rule.

To account for the indirectness of colonial rule, I controlled whether the ethnic group was under British or French colonization, and the main finding remains robust. Nevertheless, one might argue that a simple British/French dummy variable may not accurately capture the precise level of indirectness, as the extent of indirect rule varied within colonial empires. McNamee (2019) argues that the focus should be on the indirectness of colonial rule. One challenge here is that there is no direct measure of the indirectness of colonial rule at the ethnic level that I am aware of. The available measures are at the country level (e.g., see Hariri, 2012; McNamee, 2019). To utilize the country-level data, I borrow from McNamee (2019) and estimate a multi-level hierarchical model as in Robinson (2014). Robinson examines the salience of national versus ethnic identity across sub-Saharan Africa using a three-level linear probability model with random intercepts estimated at the ethnic group and country level. This multi-level hierarchical model allows us to control unchanging country-level characteristics, such as the degree of indirect colonial rule, which has been previously shown to be correlated with national versus ethnic identification.

Table AS3 displays the results obtained from Robinson (2014)’s multi-level hierarchical model, incorporating the measure of indirectness of colonial rule from (McNamee, 2019).

Panel A of Table AS3 utilizes the replication data of Robinson (2014) as reported by McNamee (2019) [[2]](#footnote-2). The results are from multi-level linear probability models from Robinson with individuals as the unit of analysis with country and ethnic group-level random intercepts. In column 1, I present estimates using only the measure of the degree of precolonial centralization. This specification reveals a negative and significant association between precolonial centralization and ethnic versus national identification (ENI). Column 2 introduces the measure of the indirectness of colonial rule at the country level from McNamee (2019). This measure of the indirectness of colonial rule is the share of customary cases recognized by colonial authorities. Consistent with McNamee’s findings, this measure demonstrates a positive and statistically significant association with the salience of ethnic identity (ENI). A similar pattern holds when using early state history (column 4), and the proportion of the population of European descent (column 5) as measures of the indirect colonial rule. In Panel B of Table AS3, the same analysis is replicated for McNamee (2019)’s cross-country sample. Notably, even controlling for the indirectness of colonial rule, precolonial centralization maintains a negative and statistically significant association with the salience of ethnic identity (ENI).

The results in Table AS3 suggest that there is potential for precolonial centralization to reverse the positive impact of colonialism on the salience of ethnic identity, emphasizing the need to consider its legacy in promoting social cohesion policies in ethnically diverse societies. One possible explanation for these findings is that colonialism introduced new institutions that overlapped with existing pre-colonial institutions in Africa. Thus, the influence of colonial institutions played a significant role in shaping the salience of ethnicity. However, over time, the temporary effects of colonial influences diminished, allowing precolonial institutions to regain their significance and increasingly shape the salience of identity. The findings align with the conclusions of Robinson (2014) and challenge the notion that the colonial legacy creates insurmountable obstacles for the development of widespread territorial nationalism in Africa. Likewise, the results are consistent with the evidence provided by Maseland (2018), suggesting that colonialism has generated a substantial yet temporary institutional shock. Moreover, these findings align with the perspective put forth by Michalopoulos and Papaioannou (2020), who argue for the limited impact of colonialism.

Nonetheless, there is one important point worth noting in passing. The results unequivocally demonstrate that precolonial political centralization is a robust predictor of ethnicity salience. However, additional analysis is warranted to ascertain whether precolonial centralization has the potential to counteract the positive influence of colonialism on the salience of ethnic identity. It is essential to investigate the specific contexts, locations, and time periods where it is likely that precolonial centralization can weaken ethnic identification by overturning the legacies of colonial policies.

# Contemporary Factors

**Political Competition** Eifert, Miguel, and Posner (2010) provide empirical evidence that supports situational theories of social identification and aligns with the perspective that ethnic identities hold significance in Africa for strategic purposes, particularly in the pursuit of political power. The authors demonstrate that ethnic identification tends to increase in the period leading up to presidential elections. Recent research suggests that the impact of political competition on the results in this paper is less substantial. Amodio, Chiovelli, and Munson (2022) find that pre-colonial centralization is associated with lower levels of political competition. Moreover, it is important to note that the evidence presented by Eifert, Miguel, and Posner (2010) focuses specifically on subnational identities, excluding national identities, and thus, the evidence by Eifert, Miguel, and Posner (2010) evidence pertains only to the salience of different subnational identities—explicitly excluding national identities. That said, however, the findings of Eifert, Miguel, and Posner (2010) - though specifically focusing on the salience of ethnicity in relation to subnational identities - could still be relevant in this case. This is because I measure ethnic identification relative to national identification, and there may be an overlap between the two measures. To address the potential influence of political competition, columns 3-5 of Table AS3 (Panel A) include controls for the proximity to an election (measured by the number of days to the nearest election) and the competitiveness of that election. The results presented in Table AS3, indicate that the main findings remain largely robust even after accounting for this contextual factor.

**Politically Relevant Ethnic Group** Posner (2004) suggests that the political importance of a cultural division is not determined by the division itself, but rather by the sizes of the groups it encompasses and their potential for political competition. Posner found that although the cultural differences between the Chewa and Tumbuka communities on both sides of the border are the same, their political significance varies greatly. He argues that this difference is due to the contrasting sizes of the Chewa and Tumbuka communities within each country in relation to the national political landscape. In Malawi, both groups are relatively large compared to the entire country, making them viable for forming political coalitions. Conversely, in Zambia, the Chewa and Tumbuka communities are small relative to the overall population and, therefore, not useful for mobilizing political support. Although size is arguably the most significant indicator of a group’s political utility, individuals belonging to any politically mobilized ethnic group—considered “politically relevant”—are expected to have a stronger inclination towards ethnic identification. To address this possibility, columns 3-5 of Table AS3 (Panel A) include a control for an indicator of politically relevant ethnic groups, and the main finding remains consistent.

Table AS3: Replications from Robinson (2014) and McNamee (2019)

 (1) (2) (3) (4) (5)

PanelA:Robinson(2014

)

’sSample

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Centralized (0-4) | -0.034\*\*\* | -0.035\*\*\* | -0.029\*\* | -0.030\*\*\* | -0.032\*\*\* |
|  | (0.011) | (0.011) | (0.012) | (0.012) | (0.011) |
| Share of Customary Court Cases |  | 0.511\*\*\* | 0.453\*\*\* |  |  |
|  |  | (0.111) | (0.119) |  |  |
| Early state history (Pre 1970) |  |  |  | 0.230\*(0.120) |  |
| Proportion of European Descendants |  |  |  |  | -3.994\*\*\*(0.683) |
| Country-Level Observations | 16 | 14 | 14 | 14 | 14 |
| Ethnic Group-Level Observations | 202 | 170 | 170 | 170 | 170 |
| Individual-Level Observations | 16,064 | 15,401 | 15,169 | 15,169 | 15,169 |
| Controls from Robinson (2014) | No | No | Yes | Yes | Yes |

PanelB: McNamee(2019

)

’sSample

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Centralized (0-4) | -0.059\*\*\* | -0.060\*\*\* | -0.058\*\*\* | -0.060\*\*\* | -0.057\*\*\* |
|  | (0.019) | (0.020) | (0.018) | (0.019) | (0.018) |
| Share of Customary Court Cases |  | 0.765\*\* | 1.297\*\*\* |  |  |
|  |  | (0.347) | (0.386) |  |  |
| Early state history (Pre 1970) |  |  |  | 0.398(0.270) |  |
| Proportion of European Descendants |  |  |  |  | -4.824\*(2.522) |
| Respondent- Level Observations | 57,382 | 57,382 | 52,507 | 52,507 | 52,507 |
| Ethnic Group-Level Observations | 204 | 204 | 202 | 202 | 202 |
| Country-Level Observations | 20 | 20 | 20 | 20 | 20 |
| Controls from McNamee (2019) | No | No | Yes | Yes | Yes |

*Notes:* The results are from multi-level linear probability models with individuals as the unit of analysis with country and ethnic group-level random intercepts. The data source is McNamee (2019). \*\*\* p*<*0.01, \*\* p*<*0.05.

**Ethnic Group in Power:** It is plausible to suggest that holding political power increases national identification, and larger ethnic groups are more likely to have access to power. Supporting this argument, Green (2020) argues that belonging to a dominant ethnic group or “Staatsvolk” and the group’s political power are significant factors in determining national identification in Africa.

Therefore, I include controls for whether an ethnic group holds political power. These include a control for group size and whether the ethnic group is in power. An ethnic group is in power if the current head of state belongs to that ethnic group. The Ethnic Power Relations (EPR) dataset provides expert coding’s of which ethnic groups hold political power. According to this dataset, a respondent is coded as belonging to an ethnic group in power if their ethnic group is identified as a “senior partner” or “junior partner” in the EPR dataset (see Robinson, 2014). Including these control variables does not alter the main results, as shown in Table AS3.

1. Panel A of Table AS3 utilizes this sample for the sake of convenience and comparability with Robinson’s (2014) study, as well as with McNamee (2019), who also employed Robinson (2014)’s sample as a robustness check. However, Panel B of Table AS3 includes McNamee (2019)’s cross-country sample to ensure that the results hold for larger samples as well. [↑](#footnote-ref-1)
2. The control variables I refer to as ‘Controls from Robinson (2014)’ in Panel A of Table AS3 are control variables from Robinson (2014) and include Competitiveness of Election, Proximity to the Nearest election (number of days), Coethnic Head of State, Ethnic Group in Power, Ln of Ethnic Group Size, Male, Formal Employment, Level of Education, Urban Residence, Ethnic Group Partition, Politically Relevant Ethnic Group, Ln of GDP Per Capita in 2005, Ethnic Fractionalization (ELF), British Colony, and Anti-Colonial War. The control variables I refer to as ‘Controls from McNamee (2019)’ in Panel B are the control variables from column 2 of Table 1 in McNamee’s (2019) study. [↑](#footnote-ref-2)