# External Threats, Linguistic Homogeneity, and Political Unification

Chen Zeng

Department of Political Science
Renmin University of China

March 15, 2022

# Introduction

Basilica di Santa Croce

Niccolò Machiavelli, a Florentine in the 16th century, argued for the unification of Italy, which didn't happen until the *Risorgimento* in 19th century, citing the following reasons:

- Florentine lifestyle
- The Italian language, esp. Tuscan dialect
- Women, god, and heroes
- External interference

# Introduction

- The English word "nation" comes from Latin *nātiōnem*, accusative of *nātiō*, equivalent to *nāscor* ("to be born") + -*tiō* ("verbal abstract noun suffix").
- In Chinese, the word for country/nation/state is "国"; it is a geopolitical concept, focusing on the territory and political power.
  - 《说文》记载："古或國同用。"
  - 根据《字源》中的考察，"或" 从口从戈，"口" 代表疆土，而 "戈" 则是兵器；后来 "或" 被假借为或许、疑惑之意，才又造出 "國" 字来。
  - 《周礼注》中又说："大曰邦，小曰國；邦之所居亦曰國。"
- Today we look at two key elements in the political unification of states—external security threats and linguistic homogeneity.

# Contents

# Political Unification

## Definition

Political unification occurs when two or more sovereign states merge into one.

- Political unification has been one of the most debated areas in the study of political science.
- Notable instances included the US (American Revolutionary War), Germany (*Sonderweg*), and Italy (*Risorgimento*).
- Does the unification happen incrementally, or is there a clear moment for such unification?
- Recent development of the EU, free trade areas, and customs unions has led political scientists to examine if there are limits to such political and economic integration. To which point do they stop being just "integration" and become unification?

# Political Unification

- An *operational* definition of political unification used in this article is that two sovereign states voluntarily merge into a federation or unitary state. Conquests/accessions are excluded.

  Unit of analysis Political unification (of a state dyad)

  Unit of observation Inter-state actions (dichotomous variable: unification, linguistic homogeneity); and individual states (continuous variable: threat levels)

- This research leverages datasets from the Correlates of War project. Qualifying states have to meet at least one of the following criteria.

  Criterion 1 Before 1920, a population of at least 500,000 and establishment of diplomatic missions at or above the rank of *chargé d'affaires* by Britain and France.

  Criterion 2 After 1920, membership in the League of Nations or United Nations *or* a population of at least 500,000 and establishment of diplomatic missions from any two major powers.

# External Security Threats

- Riker (1975) argues that states unify for security reasons; they desire protection from external threats.
- The underlying logic is that sovereignty is a valuable possession and only security-related factors will induce states to give it up voluntarily.
- With its roots in the realist tradition, this position has been supported and extended in a series of studies that address the problem theoretically, formally, and through case studies (Deudney 1995; Hechter 2000; Parent 2009; Rector 2009; Stepan 2001).
- the notion that security issues are a necessary condition for political unification has become somewhat of a conventional wisdom.

# Linguistic Homogeneity

- A common language can be posited as a necessary condition for political unification. The logic is that states will have an easier time understanding one another, and can imagine themselves as part of a larger group or nation.
- Language can be thought of as proxy for other culturally relevant factors such as nation or ethnicity. Although the distribution of languages does not correspond perfectly with nation or ethnicity, but there is a strong relationship, and language is more identifiable marker.
- To certain extent, the hypotheses of security threats and linguistic homogeneity having impact on political unification reflect the realist-constructionist divide.

# Research Hypotheses

Hypothesis 1 External threats are *necessary* for political unification.

Hypothesis 2 Linguistic homogeneity is *necessary* for political unification.

- The article takes on the form of a *falsification* probe and focuses on only *necessary* conditions for political unification.
- That is to say that security threats or linguistic homogeneity may not be *sufficient* for political unification, but the lack of either condition cannot result in the expected outcome.
- This method is prone to many methodological and substantive vulnerabilities. We will cover them in the final section, but now, let's take a look into why the author sets those two key explanatory variables.

# Cases of Unification

Table 1: Cases of Unification after 1816

| Event | Year |
|---|---|
| Piedmont – Parma | 1860 |
| Piedmont – Tuscany | 1860 |
| Piedmont – Modena | 1860 |
| Prussia – Mecklenburg Schwerin | 1867 |
| Prussia – Hesse Grand Ducal | 1867 |
| Prussia – Baden | 1871 |
| Prussia – Bavaria | 1871 |
| Prussia – Wuerttemburg | 1871 |
| France – Tunisia | 1881 |
| Japan – Korea | 1905 |
| USA – Cuba | 1906 |
| Egypt – Syria | 1958 |
| Tanganyika – Zanzibar | 1964 |
| West Germany – East Germany | 1990 |
| Yemen Arab Republic – Yemen People's Republic | 1990 |

# Within-state Comparisons

Table 2: Within-state Comparisons

| State | Threat Level (Overall) | Unification Year | Threat Level (Pre-unification) | Threat Differential |
|---|---|---|---|---|
| Egypt | 2.5 | 1958 | 11.3 | $8.8^{***}$ |
| Syria | 12.2 | 1958 | 20.3 | $8.1^{**}$ |
| Prussia | 1.4 | 1867, 1871 | 4.5 | $3.1^{**}$ |
| Piedmont | 1.0 | 1860 | 4.0 | $3^{**}$ |
| Hesse Grand Ducal | 0.2 | 1867 | 2.0 | $1.8^{***}$ |
| Mecklenburg-Schwerin | 0.2 | 1867 | 1.3 | $1.1^{**}$ |
| Baden | 0.2 | 1871 | 1.3 | $1.1^{**}$ |
| Bavaria | 0.3 | 1871 | 1.3 | $1^{*}$ |
| Wuerttemburg | 0.4 | 1871 | 1.3 | 0.9 |
| Parma | 0.0 | 1860 | 0.0 | 0.0 |
| Zanzibar | 0.0 | 1864 | 0.0 | 0.0 |
| Tanganyika | 0.0 | 1864 | 0.0 | 0.0 |
| Tuscany | 0.1 | 1860 | 0.0 | -0.1 |
| Modena | 0.3 | 1860 | 0.0 | -0.3 |
| West Germany | 1.0 | 1990 | 0.5 | -0.5 |
| Yemen Arab Republic | 1.7 | 1990 | 1.0 | -0.73 |
| East Germany | 0.9 | 1990 | 0.0 | -0.9 |
| Yemen People's Republic | 2.2 | 1990 | 0.0 | $-2.2^{*}$ |

*Note:* $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01

Table 3: Linguistic Homogeneity of Unifying States

| Event | Year | Linguistic Homogeneity |
|---|---|---|
| Piedmont – Parma | 1860 | Yes* |
| Piedmont – Tuscany | 1860 | Yes* |
| Piedmont – Modena | 1860 | Yes* |
| Prussia – Mecklenburg Schwerin | 1867 | Yes |
| Prussia – Hesse Grand Ducal | 1867 | Yes |
| Prussia – Baden | 1871 | Yes* |
| Prussia – Bavaria | 1871 | Yes* |
| Prussia – Wuerttemburg | 1871 | Yes* |
| France – Tunisia | 1881 | No |
| Japan – Korea | 1905 | No |
| USA – Cuba | 1906 | No |
| Egypt – Syria | 1958 | Yes |
| Tanganyika – Zanzibar | 1964 | Yes |
| West Germany – East Germany | 1990 | Yes |
| Yemen Arab Republic – Yemen People's Republic | 1990 | Yes |

*Note:* *Borderline case

# Reflections

In the following subsections, we let $U$ denote political unification, $S$ external security threats, and $L$ linguistic homogeneity.

- From previous analysis, the author detected 12 cases of political unification from the MID dataset, where linguistic homogeneity is present in all cases. Incorporating the Switzerland case as an outlier, we have 13 cases in total. Mathematically, we have

$$P(L|U) = \frac{12}{13} \approx 0.92 \tag{1}$$

- According to Bayes theorem, we have

$$P(U|L) = \frac{P(U)P(L|U)}{P(L)} \tag{2}$$

# Reflections

- Since $P(U)$ and $P(L)$ have the same denominator, we can simplify the equation as

$$P(U|L) = \frac{N(U)P(L|U)}{N(L)} \tag{3}$$

  whereas $N(U) = 12$, as per previous discussion.
- Only considering all 6 UN official languages, we have

$$\begin{aligned} N(L) <& N(EN)^2 + N(ES)^2 + N(FR)^2 + N(ZH)^2 + \\ & N(RU)^2 + N(AR)^2 = 2115 \end{aligned} \tag{4}$$

- Furthermore, we estimate the probability of political unification given the presence of linguistic homogeneity.

$$P(U|L) = \frac{N(U)P(L|U)}{N(L)} < \frac{12 \times 0.92}{2115} \approx 0.52\% \tag{5}$$

# Causal Inference

- Before we investigate why there is a significant gap between those two probabilities, we first consider how we should draw causal inference of political unification, taking two variables—external threats, and linguistic homogeneity—into consideration.

Table 4: Potential Outcomes of Political Unification

| | External Threats | |
|---|---|---|
| Linguistic Homogeneity | No | Yes |
| No | $Y(0,0) = ?$ | $Y(1,0) = ?$ |
| Yes | $Y(0,1) = ?$ | $Y(1,1) = ?$ |

- We first present a counterfactual analysis, and then try to relax the requirements for our evidence. As we do it, we are making stronger assumptions to draw causal inference.

# Causal Inference

## The Fundamental Problem of Causal Inference

This problem arises because we can't observe counterfactual outcomes; instead, we can only observe one of the potential outcomes.

- Research designs for causal inference, each has its own assumptions (from weak to strong)
  - Randomized Control Experiment (RCT, true randomization, assumes unit homogeneity) — usually not possible for macro-level researches in social sciences
  - Natural experiment (as-if randomization) — not applicable to this research question
  - Observational study
    - Problems: spurious relationships, omitted-variable bias, endogeneity, strong assumptions
    - Possible solutions: regression discontinuity design (RDD), difference in differences (DiD), instrumental variable (IV)...

# Selection Bias

- The problem of *external validity*—does the conclusion we draw from the sample also apply to the population?
- The sample used in this article is *not* representative—it only uses evidence available after 1816, ignoring key cases such as the American Revolutionary War. This is not a random sample.
- Omitted-variable bias is present. Other factors that may influence political unification are not controlled. Interaction cannot be taken into consideration in current research framework.
- Possible presence of confounding variables.

# Confounding factors – geographical proximity

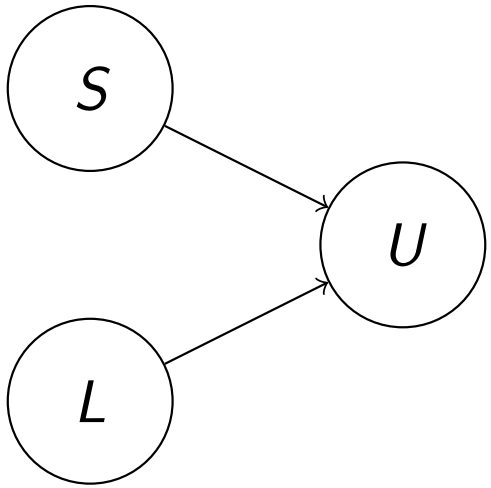


Figure 1: Causal Diagram of Griffith's Paper

# Confounding factors – geographical proximity

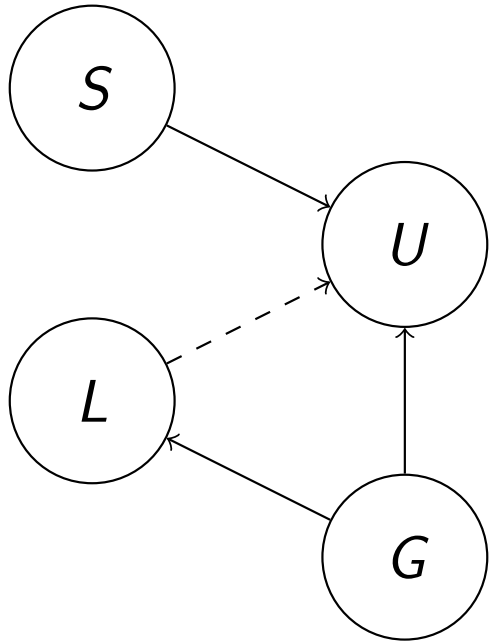


Figure 2: Causal Diagram with Spatial Proximity as A Confounder

wherein, $Cov(L, G) \neq 0$

$Cov(U, G) \neq 0$

# Alternative research design – csQCA

Table 5: Aggregated Political Unification Cases

| | Explanatory | | Outcome | |
|---|---|---|---|---|
| | S | L | U | Cases |
| 1 | No | No | ? | ? |
| 2 | No | Yes | ? | ? |
| 3 | Yes | No | ? | ? |
| 4 | Yes | Yes | ? | ? |

# Alternative research design – csQCA

Table 6: Aggregated Political Unification Cases (Hypothetical Scenario)

| | Explanatory | | Outcome | |
|---|---|---|---|---|
| | S | L | U | Cases |
| 1 | No | No | No | 800 |
| 2 | No | Yes | Yes | 50 |
| 3 | Yes | No | Yes | 100 |
| 4 | Yes | Yes | Yes | 12 |

$$\begin{aligned} U &= sL + Sl + SL \\ &= L + Sl \\ &= SL \end{aligned} \tag{6}$$

# Alternative research design – Logistic Regression

- QCA method has a few limitations.
  - It only accounts for binary explanatory variables, although we can extend csQCA to fsQCA to address this problem.
  - Outcome variable is asymmetric, i.e., for the same set of explanatory variables, we are only allowed to observe one outcome. This implies that we can't model political unification probabilistically.
  - Impossible to model interaction.
- Since the outcome variable is dichotomous, we can use logistic regression to model this problem.
- As a type of Generalized Linear Models (GLMs), logistic regression uses a link function to transform the range of the outcome variable to $[0, 1]$.

$$logit(Y) = ln(\frac{Y}{1-Y}) = \beta_0 + X\beta$$

$$\text{equivalent of } Y = \frac{e^{\beta_0 + X\beta}}{1 + e^{\beta_0 + X\beta}}, Y \sim [0, 1]$$

# Alternative research design – Logistic Regression

- Consider the following model, with $S$ and $L$ impacting the outcome variable independently.

$$\textit{logit}(P(U)) = \beta_0 + \beta_1 \times S + \beta_2 \times L \tag{7}$$

- Then consider the outlier of Switzerland, where unification occurred with external threats in the absence of linguistic homogeneity. One possibility is that there could be an interaction term.

$$\textit{logit}(P(U)) = \beta_0 + \beta_1 \times S + \beta_2 \times L + \beta_3 \times S \times L \tag{8}$$

- It means that the impact of $S$ on $\textit{logit}(P(U))$ is $\beta_1 + \beta_3 L$; and the impact of $L$ is $\beta_2 + \beta_3 S$. If $\beta_2$ is large enough, $\beta_3$ may be estimated as a negative number, meaning that in the presence of linguistic homogeneity, external security threats are of less concern when it comes to political unification.
- Are observations nested in hierarchical structures such as continents?

Thanks!