How does reported partisan attachement influence affective polarization?*

A comparative study of 25 European democracies

Tristan Muno†

November 21, 2025

You can add an abstract here.

1 Key Research Notes & Contributions

1.1 1. Conceptual & Theoretical Contribution

- **Identified Gap:** Existing studies on affective polarization either focus on **explicit partisan subsamples** or use whole samples (e.g., feeling thermometers), failing to analytically distinguish between types of attachment.
- **Analytical Distinction:** Introduces the distinction between:
  - **Explicit Partisans (Treatment,** $T = 1$): Those who explicitly report partisan attachment.
  - **Implicit Partisans (Control,** $T = 0$): Those who report **no** attachment but indicate a vote choice/intention.

*You can add acknowledgements here. Wordcount: 1387.
†University of Mannheim; Mail: tristan.muno@uni-mannheim.de

1.2 2. Empirical Mapping & Research Question

- **Initial Analysis:** Examine the **distribution of explicit and implicit partisans** across 25 EU countries.
- **Figure 1:** Displays the cross-national distribution, establishing **substantial cross-national variation**.
- **Core Question: Does the explicit vs. implicit partisan distinction *matter* for behavior?** (I.e., what is its causal effect?)

1.3 3. Causal & Methodological Framework

- **Causal Objective:** Estimate the **conditional average causal treatment effect (CATE)** of being an Explicit Partisan ($T = 1$).
- **Outcomes (QoIs):**
  - Affective Polarization ($Y_C - Y_O$)
  - In-group Favoritism ($Y_C - Y_{Non}$)
  - Out-group Derogation ($Y_{Non} - Y_O$)
- **Methodology:** Multilevel Regression Adjustment (Bayesian *brms*) using a **single, combined model** for all token allocations, including:
  - **Fixed Effects:** Treatment (T), Recipient Type (R: Co/Out/Non), and the crucial $\mathbf{T} \times \mathbf{R}$ interaction (to directly estimate QoIs).
  - **Adjustment:** Control for key covariates (X) to satisfy the Conditional Independence Assumption (CIA).

– **Structure:** Accounts for nesting via random effects (Rounds → Respondents → Parties → Countries), specifically including **random slopes for** T at the party level ($\mathbf{1 + T \mid party\_id}$) for heterogeneity control.

1.4 4. Results & Findings Presentation

- **Figure 2 (Pooled AMCE):** Presents the **Average Marginal Component Effect (AMCE)** for all three QoIs across both game types (Dictator and Trust Game), summarizing the overall causal contribution of explicit attachment across 25 countries.
- **Heterogeneity Analysis:** Investigates substantial **country-level variation** in the causal effect.
- **Figures 3, 4, 5 (Country-Specific Effects):** Disaggregate the treatment effect for Affective Polarization, In-group Favoritism, and Out-group Derogation by country and game type. These figures utilize the estimated **random slopes and intercepts** from the multilevel model to show where the explicit/implicit distinction matters most.

2 Introduction

Recent advancements in the study of affective polarization point at the importance of clearly defined conceptualization. The seminal work of Iyengar introduced the concept of affective polarization primarily as a contrast to ideological polarization in the debate on political polarization in the US. Since then, a lot of research has been dedicated to

studying the causes (why is it a thing), concequences (what does it do or more normatively, why should this be considered a problem or not), and dynamics (how does it change) of affective polarization beyond the American case.

This lead to a sharp increase in publications on affective polarization. Along the way the discipline matured regarding its conceptualization, acknowledging the inherent multi-dimensionality (or at least ambiguity) of the term affective polarization. As XYZ point out, "[t]he terms"affective polarization", "partisan polarization", or "partisan affective polarization" have long been used interchangeably in the literature". Similarly, empirical measures of affective polarization are often applied without reflection regarding the conceptual and operationalization choices (make sure its not too much paraphrasing). Some studies focus on partisan subsamples, others include the whole electorate (sample), but little research has investigated whether there are systematic differences.

This study addresses this gap by introducing the analytical distinction between *explicit* partisans and *implicit* partisans, investigating their relative distributions, and studying their differences in affective polarization behavior. By explicit partisans we mean respondents who report a subjective attachement to a political party. We term respondents indicating they do not feel personally attached to any political party, but that reported a vote intention or choice, implicit partisans. Logically, this corresponds to previous studies' differentiating between partisans and supporters, but we believe it to be useful to refer to both as partisans, but two different types. (It may better account for reality's complex nature, overcoming the binary heuristic often applied in social identity theory approaches by overcoming the simple binary distinction of either you are in a group or not, though admittably, we reduce complexities of multi-party systems at the other end when averaging across out-partisans).

Empirically, we aim to isolate the average treatment effect of explicit partisanship

on affective polarization. Thus, we conceptually think of reported partisan attachement as the (heterogeneous) treatment, and affective polarization respecitvely its subcomponents, ingroup favoritism and outgroup derogation, as the outcome(s). Using hierarchical bayesian modelling and data from 25 European countries, we focus on affective polarization in multi-party systems.

Our results reveal …

These findings have implications for affective polarization research: When (conceptualizing) and measuring affective polarization, scholars should dedicate some thought onto whether the population of interest, onto which they aim to generalize their findings, consists of *explicit* partisans or the whole electorate of explicit and implicit partisans. Depending on the specific case, these two sets are notably different in terms of size, and drawing inferences based on (explicit) partisan subsamples alone might yield biased conclusions about the overall state of affective polarization in a given society.

3 Theory

We begin by outlining existing conceptualization and measurements in affective polarization literature, with a focus on studies dedicated to European multi-party systems.

Afterwards, we introduce our analytical distinction and present our argument why inferences based on explicit partisan subsamples encompasses potential for biases in more detail.

Finally, we synthesize existing literature on partisanship and affective polarization to formulate hypothesis about potential differences between explicit and implicit partisans regarding their affective polarization behavior.

Hypotheses:

H_1: Explicit partisans exhibit higher levels of affective polarization than implicit partisans.

H_2: Explicit partisans exhibit higher levels of ingroup favoritism than implicit partisans.

H_3: Explicit partisans exhibit higher levels of outgroup derogation than implicit partisans.

Note that these hypotheses are logically dependent: As ingroup favoritsm and outgroup derogation are two subcomponents that jointly make up affective polarization, if H_1 were true, either H_2 or H_3 or both have to be true by extension.

H_2: Explicit partisans exhibit both more ingroup favoritsm and more outgroup derogation than implicit partisans.

depending on literature, additional comparative hypothesis

4 Research design

In this section, we situate our measurement approach of affective polarization, present the used data and out empirical strategie to approximate an isolation of the causal effect of *explicit* partisanship on our outcomes of interest: affective polarization, ingroup favoritism and outgroup derogation.

In causal terminology, explicit partisans form our treatment group ($T = 1$), with implicit partisans constituting our control group ($T = 0$).[1]

[1]We omit the prefix quasi- for readability when referring to treatments, controls, and causal terminology. This choice is purely stylistic and does not imply that our design is experimental.

The underlying idea is to study the effect of explicit subjective group attachement compared to implicit group attachement on affective polarization.

The effect of the partisan-group-relative-variable is conditional on the type of partisanship. The partisan-group-relative-variable is itself a function of assigned partisan ingroup of respondents (by attachement for explicit and by vote intention/choice for implicit) and the displayed party affiliation of the conjoint profile.

4.1 Data

We analyze data from Hahm, Hilpert, and König (2024) …

4.2 Independent variables

We code…

Partisanship

$T = 1$ -> explicit partisanship $T = 0$ -> implicit partisanship

Conjoint stimuli Z

Partisan Group-relative variable (Co-, Out- or Non-partisan) P_g

4.3 Dependent variable

Token allocated in the experiment

Y

4.4 Control variables

4.5 Causal model

Partianship -> Conjoint_Profiles -> Tokens

4.6 Statistical model

Let Y_{ripc} be the number of tokens allocated in round r by respondent i of party p in country c.

Then,

$$Y_{ripc} = \beta_1 * P_G + \beta_2 * T + \beta_3 * P_G * T$$

tbd

nested random intercepts at i, p, and c and random effects for p and c

to account for party level variation in both the general extend of token spending patterns and the specific effect of explicitly feeling attached. and for comparative analyses of country differences in the effects of T

4.7 Quantities of interest

4.7.1 Affective polarization

Measure: - Explicit partisans: - (Expected number in tokens allocated | coplayer is copartisan & explicit partisanship) - (Expected number in tokens allocated | coplayer is outpartisan & explicit partisanship) - Implicit partisans: - (Expected number in tokens

allocated | coplayer is copartisan & implicit partisanship) - (Expected number in tokens allocated | coplayer is outpartisan & implicit partisanship)

CATE = AffPol_Explicit - AffPol_Implicit

4.7.2 Ingroup favoritms

Measure: - Explicit partisans: - (Expected number in tokens allocated | coplayer is copartisan & explicit partisanship) - (Expected number in tokens allocated | coplayer is nonpartisan (control) & explicit partisanship) - Implicit partisans: - (Expected number in tokens allocated | coplayer is copartisan & implicit partisanship) - (Expected number in tokens allocated | coplayer is nonpartisan (control) & implicit partisanship)

CATE = InFav_Explicit - InFav_Implicit

4.7.3 Outgroup derogation

Measure: - Explicit partisans: - (Expected number in tokens allocated | coplayer is nonpartisan (control) & explicit partisanship) - (Expected number in tokens allocated | coplayer is outpartisan & explicit partisanship) - Implicit partisans: - (Expected number in tokens allocated | coplayer is nonpartisan (control) & implicit partisanship) - (Expected number in tokens allocated | coplayer is outpartisan & implicit partisanship)

CATE = OutDer_Explicit - OutDer_Implicit

5 Empirical analysis

We present our findings in three stages. First, we present the distributions of explicit and implicit partisanship across our included country samples. Next, we present the overall

average treatment effects of explicit partisanship on our outcomes of interest. Finally, we present cross-country variation of these pooled results to place our results in a comparative perspective.

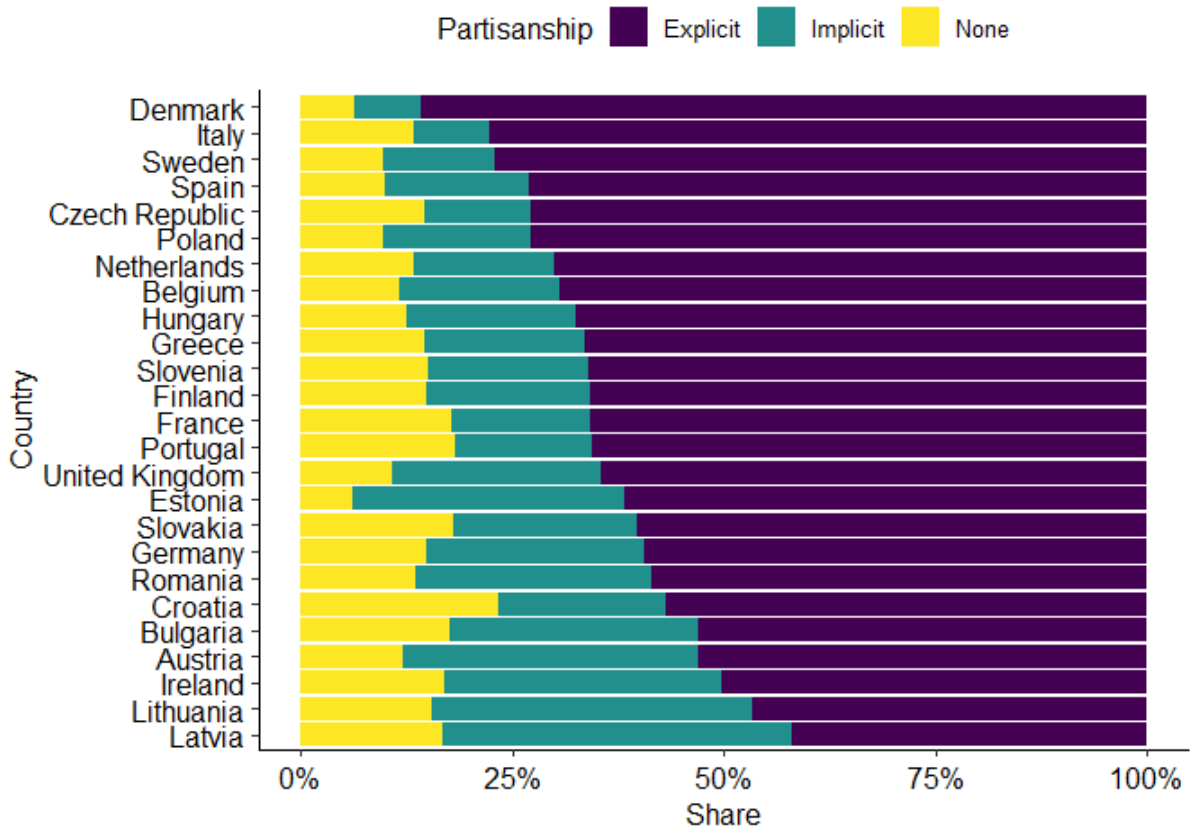

Figure 1

6 Conclusion

7 References

Hahm, Hyeonho, David Hilpert, and Thomas König. 2024. "Divided We Unite: The Nature of Partyism and the Role of Coalition Partnership in Europe." *American Political Science Review* 118 (1): 69–87.

8 Appendix