How does partisan type influence affective polarization?*

A comparative study of 25 European democracies

Tristan Muno[†]

November 30, 2025

This study introduces the concept of partisan *type*, which distinguishies *explicit* partisans – respondents who explicitly report attachement to a political party – from *implicit* partisans, namely respondents who explicitly report no attachement, but still express a partisan preference in standard voting items. Existing studies on affective polarization often use measures that restrict analyses to either type trying to infer the state of affective polarization in a given context. This study explores systematic differences between explicit and implicit partisans regarding their affective polarization behavior, reanalyzing experimental data from 25 European multi-party systems. Results show that 1) shares of explicit partisans vary from 45% to 85%, 2) implicit partisans show almost identical levels of affective polarization, 3) lower levels of IF and OD 4) this holds across countries. The implications are that scholars need to carefully consider what their population of interest is when measuring affective polarization, as context-specific differences in the relative distributions of partisan types may yield biased inferences about the state or dynamic of polarization in a given society.

1 Introduction

How does partisan type influence affective polariztion? While the past decade witnessed a sharp increase in publications dedicated to study the causes, consequences, and dynamics of affective polarization, there still remains much ambiguity when it comes to conceptualization and measurement. Notably, some conceptual or measurement approaches limit their scope to only a partisan subset of the sample, while other strategies cover a wider

*You can add acknowledgements here. Wordcount: 3561.

[†]University of Mannheim; Mail: tristan.muno@uni-mannheim.de

range of the sampled electorate. Nevertheless, the goal is often times identical: infer about the state or change of affective polarization in a given country's society. If however such inferences are based on partisan subsets, inferring statements about that subset to the whole population rest on assumptions about the distribution of that subset in the population. In other words, only if the partisan subset a) is large enough to constitute a (sufficiently) large part of the population or b) is representative of that population, one can validly infer from the subsample to the population.

This study aims to address this gap and contribute to the affective polarization literature in multiple ways. First, I introduce the concept of *partisan type* to analytically distinguish between *explicit* and *implicit partisans*, enabling affective polarization scholars to better clarify what their sample of study and population of interest is. Second, I empirically investigate the relative distributions of partisan types across 25 European multi-party system, finding cross-national variation in the relative shares but surprising consistencies regarding the overall numbers of partisans. Third, I employ a causal identification strategy using hierarchical bayesian modelling to approach an answer to the question: How would a hypothetical intervention changing a respondent's partisan type impact her affective polarization behavior, holding all other covariates constant? Finally, I put things in a comparative perspective by studying cross-national variation in the hypothesized relationship between partisan type and affective polarization outcomes.

From a theoretical perspective, affective polarization emphasizes individuals' identities, their subjective feelings of belonging to a perceived group and the corresponding, emotional, *affective*, mechanisms that subconsciously influence human intergroup behaviors. Hence, through this social identity perspective, we expect *explicit* partisan types to show stronger levels of affective polarization. The rationale is that explicit reporting of feelt attachement corresponds to stronger group identites, Paragraph on research design (causal strategy)

Paragraph on results

Paragraph on implications (conditional on results)

If implicit partisans polarized too -> suggests social identity mechanism despite clear rejection of subjective attachement -> identity complex

Differences -> approaches like Hahm, Hilpert, and König (2024) less suited to infer about state of polarization in a given country context

2 Theory

2.1 Lit Review

Aff Pol -> What is it and where it comes from (literature) -> polarization as static presence of bi- or multi-modal distributions -> polarization as dynamic increase both make sense, I focus on first

The concept of affective polarization, losely defined as the tendency to favor one's own party and disliking other parties, has received considerable attention following Iyengar, Sood, and Lelkes (2012)'s seminal paper. Since then, the social identity framework of political parties has been extended beyond the American case. According to classical social identity approaches,

Aff Pol -> what are causes, consequences, and dynamics (state of research)

aff pol closely related to partisanship and identity, but what is partisanship, what is identity? -> literature on partisanship -> introducing partisan type

Aff Pol Measurement/Conceptual ambiguities (problems) -> often no clear definition of partisan type

Aff Pol ambiguities in Multiparty Systems/Comparative analyses specifically and recent trends (problems in comparative research) -> not clearly defining partisan type that is being studied poses comparative risks -> e.g. if in country A 70% explicit partisans but country B 50% explicit partisans -> using explicit partisan subsamples alone, inference over state of polarization in country A and country B differing (worse for latter / prone to bias) -> if, e.g., explicit partisans are systematically higher polarized -> overestimation of country Bs polarization level

2.2 Argument

Concept of partisan type, distinguishing implicit and explicit partisan types, useful for addressing abiguities regarding subsamples and populations of interest.

Social identity theory: explicit types -> reported attachement (subjective belonging) -> indicates group identity -> higher ingroup outgroup mechanism

implicit ones: explicitly report no attachement -> no direct/conscious group attachement -> expressive group preference nevertheless through vote variables -> less polarization

-> HTs 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

3 Research design

- Data and experimental setup -> waves and survey items -> hierarchical data structure
- Variables -> DAG -> tripartite structure -> recapture hierarchical structure
- Causal Assumptions
- Model -> recapture hierarchical structure -> cross level interaction
- Control Variables? Priors
- Quantities of Interest: CATE AP, IG, OD

This section outlines how we measure affective polarization and its subcomponents, the data we use, and out empirical strategy to estimate the effect of explicit partisanship on our outcomes of interest: affective polarization, ingroup favoritism, and outgroup derogation.

We conceptually frame explicit partisanship as a treatment and implicit partisanship as a control. That is, respondents reporting a subjective attachement to a political party

($T_i = 1$) form the treated group, whereas respondents expressing vote choice without attachement ($T_i = 0) form the comparison group.[1]

The underlying research question is: Does explicit partisan attachement amplify affective polarization relative to implicit partisan attachement?

To approximate causal inference from observational variation, we exploit the following features of the data:

1. Tripartite structure of partisan and profile relationships. Each respondent i is characterized by their *partisan type* (T_i) and a *partisan anchor* (A_i), which is either the party they feel attached to (explicit partisans, $T_i = 1$) or intend to vote for (implicit partisans, $T_i = 0$). Each conjoint profile presents a party cue (Z_{ri}^{party}), which is compared to the respondent's anchor to classify the profile as co-partisan, out-partisan, or neutral ($R_{ri} \in Co, Out, None$). This structure separates treatment assignment, i.e., partisan type (T_i), respondent identity, i.e., partisan anchor (A_i), and profile-level variation ($R_r i$).
2. Randomized conjoint profiles. The profiles' attributes – including all non-partusab cues Z_{ri}^{other} – are randomly assigned. This ensures that observed differences in token allocations across $R_r i$ are attributable to profile characteristics rather than respondent selection.
3. Hierarchical variation. Respondents are nested within partisan anchors, which are themselves nested within countries. This allows for multilevel modeling that captures heterogeneity at the respondent, party, and country level.

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

[1] For readability, we omit the prefix "quasi-" when referring to treatment and control, but this does not imply random assignement. Our design remains observational.

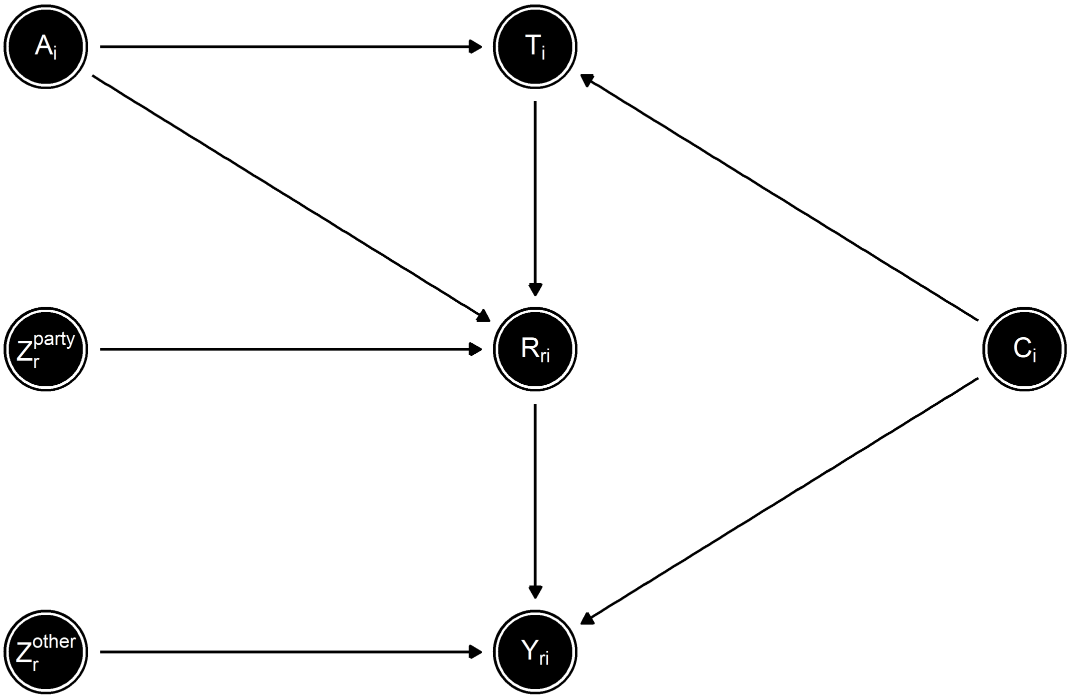

Figure 1: Directed acyclic graph of the causal data-generating process. Respondents have a partisan anchor A_i, representing the party they feel attached to (explicit partisans) or intend to vote for (implicit partisans). Partisan type T_i (explicit vs. implicit) is only defined for respondents with a partisan anchor and is therefore a child node of A_i. Each conjoint profile r presents randomized attributes: a partisan cue Z_r^{party} and other cues Z_r^{other}. The partisan relationship category $R_{ri} = f(A_i, Z_r^{party})$ determines whether the profile is interpreted as a co-partisan, out-partisan, or neutral. Token allocations Y_{ri} are affected by both R_{ri} and Z_r^{other}, with the effect of Rri theorized to depend on T_i. Because T_i is observational, C_i denotes potential confounders of both T_i and Y_{ri}, highlighting the assumptions required for causal interpretation.

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]

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.

Respondents vary along two analytically distinct dimensions. First, the partisan type (T_i) distinguishes respondents who report subjective partisan attachment (explicit partisans, $T_i = 1$) from those who report a vote intention or choice in the absence of attachment (implicit partisans, $T_i = 0$). Second, each respondent is assigned a partisan anchor (A_i), defined as the party to which they report either attachment or vote intention. In the conjoint tasks, the party cue of profile j is compared to anchor A_i to generate a partisan relationship category ($R_{ij} \in \{Co, Out, Non\}$). This tripartite structure cleanly separates treatment status, party identity, and respondent–profile relationships, and it allows us to estimate the conditional effect of explicit partisan attachment across parties and countries using a hierarchical model with random effects for respondents, parties, and countries.

3.1 Data

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

3.2 Variables

Following our theoretical framework outlined before, we now operationalize our main variables.

Our main variable of interest is $T_i = \{1, 0\}$, which denotes the *partisan type* of respondent i. Values of $T_i = 1$ indicate explicit partisan type, i.e., respondents who reported a feeling of attachement to any political party. $T_i = 0$ marks implicit partisan types: respondents who actively reported not feeling attached to a political party, but expressed a party preference through a vote choice or intention variable.

Next, each respondent is assigned a *partisan anchor* (A_i), namely the party respondents indicated to feel attached to or has (intended to) voted for.

We use $Z_r i$ to denote the conjoint profile's cue presented in round r to respondent i. These cues include dimension party, age, class, gender, religion, nationality and in case of the dictator game an EU attachement. Z_{ri}^{party} marks the partisan cue, Z_{ri}^{other} all other conjoint cues.

Given we are interested in the partisan effects, we limit observations to conjoint profiles with the nationality attribute fixed to that of the respective country (i.e., "co-national" from respondents' point of view), since there was never a partisan cue for outnationals to exclude implausible scenarios.

Figure X displays the causal data-generating structure. Respondents possess a partisan anchor A, defined as the party they feel close to (explicit partisans) or intend to vote for (implicit partisans). Partisan type T (explicit vs. implicit) is only defined for respondents with a partisan anchor and is therefore a child node of A. The conjoint design randomizes all profile attributes. The partisan cue \$Z^{party} combines with A to produce the partisan relationship category $R = f(A, Z^{party})$, which determines whether the profile is interpreted as co-partisan, out-partisan, or neutral. Other randomized profile cues Z^{other} affect the outcome directly. Token allocations Y are affected by both R and Z^{other}, with the effect of R theorized to be conditional on T. Because T is observational, we cannot assume conditional independence a priori, hence C denotes potential confounders of both T and Y.

To interpret the path $T \to R \to Y$ as the conditional causal effect of T, we must establish the following assumptions:

SUTVA -> Treatment outcomes are independent of treatment assignment to other units Y_i(1), Y_i(0) Conditional Independence -> No confounders influencing T and Y Covariate Balance -> $E[X \mid T = 1] = E[X \mid T = 0]$ Comparability of T=1 and T=0 Common Support -> for every respondent, there is a positive probability of being assigned to both the treatment and control groups (fulfilled by subset to respondents with partisan anchor)

3.3 Independent variables

We code...

Partisanship

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

3.4 Dependent variable

Token allocated in the experiment

Y

3.5 Control variables

Given the observational and non-random assignmend of T_i, including any potential confounding covariates is crucial to ensure comparability across the treatment levels $T_i \in 1, 0$ and to block backdoor paths that jointly influence partisan type T and token allocations Y.

Although the goal of completely including all covariates is untestable, we try to include as many relevant covariates as possible, given the numerous studies on partisanship.

Given the observational nature of the partisan type variable (Ti), we employ a comprehensive set of control variables (Ci) to fulfill the Conditional Independence Assumption and minimize confounding. This set includes standard demographic and socioeconomic characteristics (e.g., age, gender, education, economic perception), core ideological positions (Left-Right, EU position, nativism), and specific psychological traits. Critically, we include a detailed battery of items measuring Anti-System and Populist Attitudes alongside general political efficacy. This rigorous control strategy ensures we isolate the treatment effect—the conditioning impact of explicit subjective attachment on affective polarization—from alternative explanations driven by generalized political engagement or anti-establishment resentment common to the implicit partisan group.

Because the inclusion of that many covariates reduces the amount of observations due to noise and missing values in the data, I estimate a base model with only experimental controls and a covariate model, which includes the above mentioned variables. The purpose is to ensure substantive interpretation is not sensitive on differing datasets depending on the covariate specification. On the other hand, the inclusion of this many covariates aims at blocking any backdoor paths between Z and Y, accounting for selection effects of Z and consequently improving the comparability between treatment groups.

I include all covariates for which the number of NAs is < 10000.

3.6 Causal model

Partisan Type -> Partisan Anchor -> Relationship Category -> Allocation

3.7 Statistical model

The unit of analysis is the conjoint round r, i.e., the evaluation of a single conjoint profile within one game. Each profile evaluation r is nested within respondent i, who is nested within a partisan anchor a (the respondent's party of attachment or vote intention), which is nested within country c. This yields a four-level hierarchical data structure:

Level 1: profile evaluations $(r) \rightarrow$

Level 2: respondents $(i) \rightarrow$

Level 3: partisan anchors $(a) \rightarrow$

Level 4: countries (c).

Respondents vary along two analytically distinct dimensions. First, partisan type T_i differentiates respondents who report subjective partisan attachment (explicit partisans, $T_i = 1$) from those who report a vote intention or vote choice without attachment (implicit partisans, $T_i = 0$). Second, each respondent is assigned a partisan anchor A_i, defined as the specific party to which they report either attachment (explicit partisans) or vote intention (implicit partisans).

In each conjoint round r the party cue displayed in the profile is compared to the respondent's partisan anchor A_i to generate a partisan relationship category $R_{ri} \in \{\text{Co}, \text{Out}, \text{None}\}$. This tripartite structure—partisan type, partisan anchor, and respondent–profile relationship—separates (1) treatment status, (2) party identity, and (3) relational meaning of the cue and allows estimation of how explicit partisan attachment conditions sensitivity to partisan cues across respondents, parties, and countries.

Let Y_{riac} denote the number of tokens allocated by respondent i (nested in partisan anchor a and country c) in conjoint round r. To estimate how explicit partisan type ($T_i = 1$) modifies the effect of the partisan relationship cue (R_{ri}) we specify the following multilevel model:

$$\begin{aligned}
Y_{riac} = \beta_0 + \beta_1 T_i + \beta_2 R_{ri} + \beta_3 (T_i \times R_{ri}) + \beta_4 Z_{ri}^{\text{other}} + \beta_5 C_i \\
+ u_{i0} \\
+ v_{a0} + v_{aT} T_i + v_{aR} R_{ri} \\
+ w_{c0} + w_{cT} T_i + w_{cR} R_{ri} \\
+ \varepsilon_{riac}.
\end{aligned}$$

Key differences to the earlier draft are that respondents enter the model only with a random intercept u_{i0} (no respondent-level random slope for R_{ri}). Initially we attempted to estimate a respondent-level random slope for R_{ri}, but within-respondent variation in the partisan-relationship variable was too limited to support a reliable estimate; consequently the final estimated specification contains respondent intercepts only and random slopes for R_{ri} (and T_i) at the partisan-anchor and country levels.

Random effects are modeled as multivariate normal with correlated intercepts and slopes at the partisan-anchor and country levels:

$$u_{i0} \sim \mathcal{N}(0, \sigma_{u0}^2),$$

$$\begin{pmatrix} v_{a0} \\ v_{aT} \\ v_{aR} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma_a \right),$$

$$\begin{pmatrix} w_{c0} \\ w_{cT} \\ w_{cR} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma_c \right),$$

$$\varepsilon_{riac} \sim \mathcal{N}(0, \sigma^2).$$

Both Σ_a and Σ_c are full 3×3 covariance matrices, so intercepts and slopes are estimated as correlated at the partisan-anchor and country levels. This mirrors the model code where the random terms for country and for partisan-anchor-within-country are specified as correlated intercepts and slopes.

3.7.1 Model structure and justification

Cross-level interaction of substantive interest.

The main theoretical question is whether explicit partisan attachment ($T_i = 1$) alters responsiveness to co-, out-, and neutral partisan cues (R_{ri}). The fixed-effect interaction $\beta_3(T_i \times R_{ri})$ estimates this conditional effect at the population (grand-mean) level.

Random slopes for higher-level variation in cue sensitivity.

The effect of the Level-1 cue R_{ri} is allowed to vary across partisan anchors and across countries (random slopes v_{aR} and w_{cR}), which captures heterogeneity in cue sensitivity that is plausibly organized by party identity and national political context. Similarly, the

Level-2 treatment T_i is allowed to vary across anchors and countries (random slopes v_{aT} and w_{cT}), reflecting that the meaning and strength of explicit partisanship differ across parties and national settings.

Respondent-level variation and identification.

Respondents vary in baseline generosity and provide multiple conjoint rounds; this is modeled with a respondent random intercept u_{i0}. We did not include a respondent-level random slope for R_{ri} because there is insufficient within-respondent variation in the relationship variable to estimate it reliably. Put differently, many respondents do not experience enough co/out/none combinations across rounds for stable respondent-specific slope estimates. Omitting a respondent-level slope implies that heterogeneity in cue sensitivity is captured at the partisan-anchor and country levels rather than at the individual level.

Implications of omitting respondent-level random slopes.

Omitting u_{iR} means the model does not estimate idiosyncratic respondent-level deviations from the average relationship effect. If relevant individual-level heterogeneity were present but unmodeled, the estimated higher-level variances (in Σ_a and Σ_c) or residual variance could absorb some of that heterogeneity. Given the data limitation (low within-respondent variation), estimating u_{iR} would be poorly identified and could destabilize inference; the chosen specification therefore balances theoretical desiderata and empirical identifiability.

Correlated random effects and partial pooling.

Correlated intercepts and slopes at the anchor and country levels allow partial pooling of treatment and cue effects, borrowing strength across units while permitting systematic contextual differences. This stabilizes estimates of heterogeneous treatment effects and aligns with the hypothesis that both party identity and national context shape partisan cue effects.

Estimation.

Models were estimated in a Bayesian framework (via *brms*), which propagates uncertainty through the multilevel interaction structure and regularizes the maximal random-effects specification through priors.

In sum, the estimated model aligns the theoretical structure of partisan cue processing with a multilevel design that
(i) treats explicit partisanship as a Level-2 treatment whose effect can vary across parties and countries,
(ii) captures heterogeneity in cue sensitivity at the partisan-anchor and country levels, and
(iii) models respondent-level repeated measures via random intercepts while avoiding poorly identified respondent-level random slopes given the observed within-respondent variation.

3.7.2 Prior specification

- **Approach 1 — Very Weakly Informative Priors (`normal(0, 5)`)**

This approach aims to be conservative, broad, and minimally informative.

Rationale

- Token allocations range from **0–10**, so any plausible effect of a predictor—main or interaction—must lie within approximately **−10 to +10**.

- A `Normal(0, 5)` prior places ~95% of its probability mass within this range (±9.8).

- This reflects no strong prior beliefs about the size or direction of effects while permitting large shifts if supported by the data.

Specification

- **Fixed effects (slopes & interactions):**
    - (_k Normal(0, 5))
- **Intercept:**
    - (Intercept Normal(5, 5))

      (midpoint of 0–10 with substantial uncertainty)
- **Random effect SDs:**
    - (_{RE} Student-t(3, 0, 5))
- **Residual SD:**
    - (Student-t(3, 0, 5))

Characteristics

- Extremely permissive (barely regularizing).

- Useful when the goal is **maximal conservatism** and **minimal prior influence**.

- May yield wider posteriors, more computation difficulty, and greater risk of overfitting with many random effects.

- **Approach 2 — Weakly Informative, Regularizing Priors (recommended)**

Aimed at *constrain-but-not-restrict* modeling: improves stability, avoids implausibly large effects, yet remains substantively uninformative.

Rationale

- Shifts of **4+ tokens** (40% of the scale) are already large in practice.

- `Normal(0, 2)` limits 95% prior mass to about ±4, which remains wide enough while avoiding unrealistic extremes.

- More appropriate for complex hierarchical models with many random slopes.

Specification

- **Fixed effects (slopes & interactions):**
  - (_k Normal(0, 2))
- **Intercept:**
  - (Intercept Normal(5, 2))
- **Random effect SDs:**
  - (_{RE} Student-t(3, 0, 1))
    or a half-normal(0,1) prior
- **Residual SD:**
  - (Student-t(3, 0, 2))

Characteristics

- Still weakly informative, but more realistic given the bounded 0–10 outcome scale.

- Improves convergence and shrinkage in large multilevel structures.

- Reduces the probability of implausible prior predictive draws (e.g., (Y < -10)).

3.8 Quantities of interest

In the following sections we define our substantive quantities of interest, affective polarization and its subcomponents, ingroup favoritism and outgroup derogation. Our substantive interest concerns the differences in these three measures between implicit and explicit partisan types. In the causal framing, we want to unravel the effect of a hypothetical intervention: If we hold all covariates constant and only change T from 0 to 1, how does this impact our quantities of interest?

3.8.1 Affective polarization

More precisely, we define affective polarization as the difference in expected token allocated in the conjoint games to co-partisans and out-partisans. This approach follows (?) Wagner (or Reiljan) in that it averages across all outgroups, which is a simplification given we are dealing with multi-party systems. Nevertheless, this (CHECK) yields a conservative measure, and allows for cross-national comparisons, as the number of parties in a given national party system varies, and so does the potential number of political outgroups.

Formally, we define affective polarization AP as:

$$AP = E(Y \mid R = Co) - E(Y \mid R = Out)$$

These conditional expectations are further refined to capture the effect of explicit attachement:

$$\begin{aligned} CATE_{AP} &= AP_{T=1} - AP_{T=0} \\ &= [E(Y \mid R = Co, T = 1) - E(Y \mid R = Out, T = 1)] \\ &\quad - [E(Y \mid R = Co, T = 0) - E(Y \mid R = Out, T = 0)] \end{aligned}$$

With the formulated hypothesis above, we posit that $CATE_{AP} > 0$: We expect respondents with explicit partisan type ($T = 1$) to exhibit higher levels of affective polarization than implicit types ($T = 0$).

3.8.2 Ingroup favoritms

Do disentangle affective polarization into ingroup favoritism and outgroup derogation, we utilize that the conjoint setup included a neutral control group, in which no partisan cue was displayed ($R = None$). This allows as to estimate the effects of a copartisan cue ($R = Co$) relative to that neutral baseline, conditional on partisan type T.

Formally, we define ingroup favoritism IF as:

$$IF = E(Y \mid R = Co) - E(Y \mid R = None)$$

further refining this equation with respect to partisan type T, we define our causal estimand as:

$$
\begin{aligned}
CATE_{IF} = IF_{T=1} - IF_{T=0} \\
&= [E(Y \mid R = Co, T = 1) - E(Y \mid R = None, T = 1)] \\
&\quad - [E(Y \mid R = Co, T = 0) - E(Y \mid R = None, T = 0)]
\end{aligned}
$$

3.8.3 Outgroup derogation

Similarly, the neutral control group ($R = None$) lets us estimate outgroup derogation as the average number in tokens withheld from outpartisans compared to the baseline condition.

Hence, we formalize outgroup derogation OR as

$$OR = E(Y \mid R = None) - E(Y \mid R = Out)$$

the subsequent causal estimand is thus

$$
\begin{aligned}
CATE_{OR} = OR_{T=1} - OR_{T=0} \\
&= [E(Y \mid R = None, T = 1) - E(Y \mid R = Out, T = 1)] \\
&\quad - [E(Y \mid R = None, T = 0) - E(Y \mid R = Out, T = 0)]
\end{aligned}
$$

By focusing on expected values, or marginal means, we also ensure our substantive results are not dependent on the choice of reference categories (Leeper, Hobolt, and Tilley 2020).

4 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.

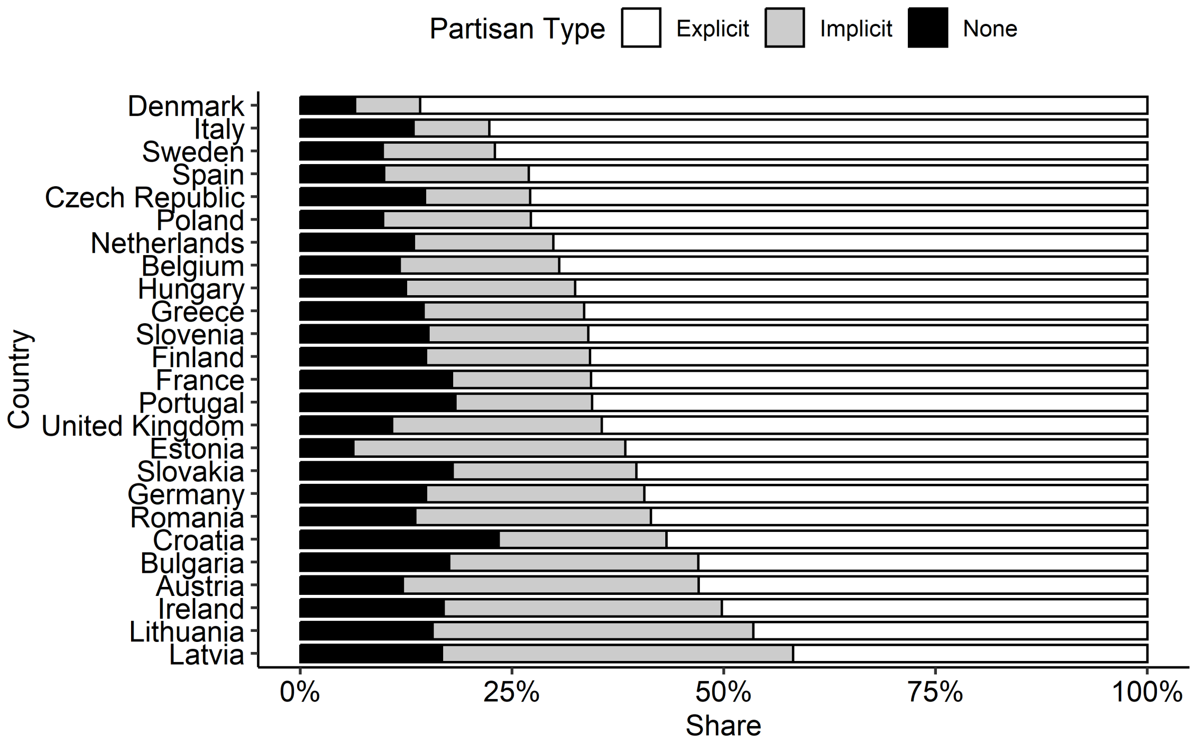

Figure 2: Distribution of Explicit, Implicit, and Non-Partisans across 25 Countries

4.1 Pooled analysis

Figure 3 shows the results for the dictator game

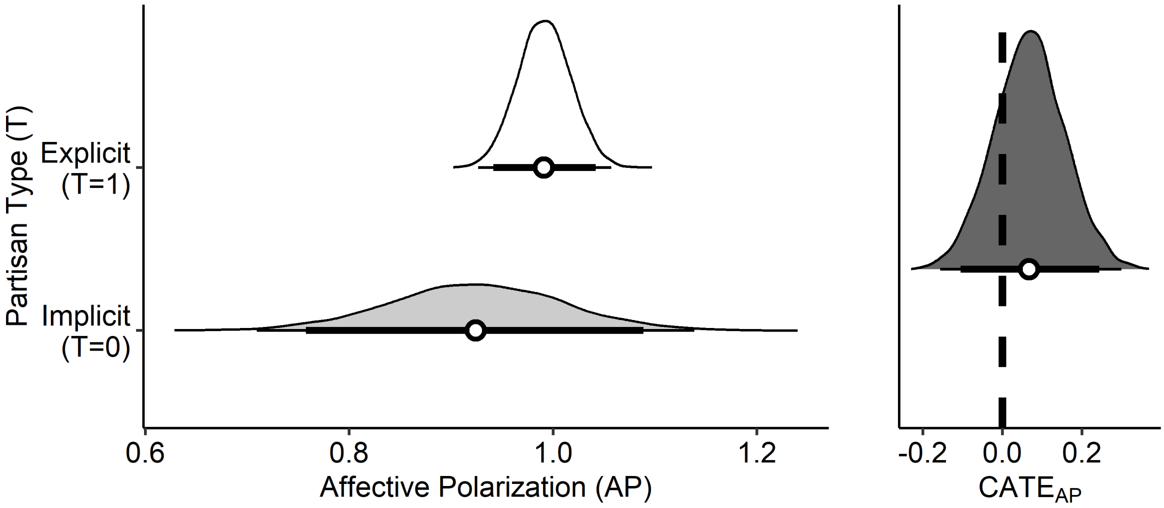

Figure 3: A title

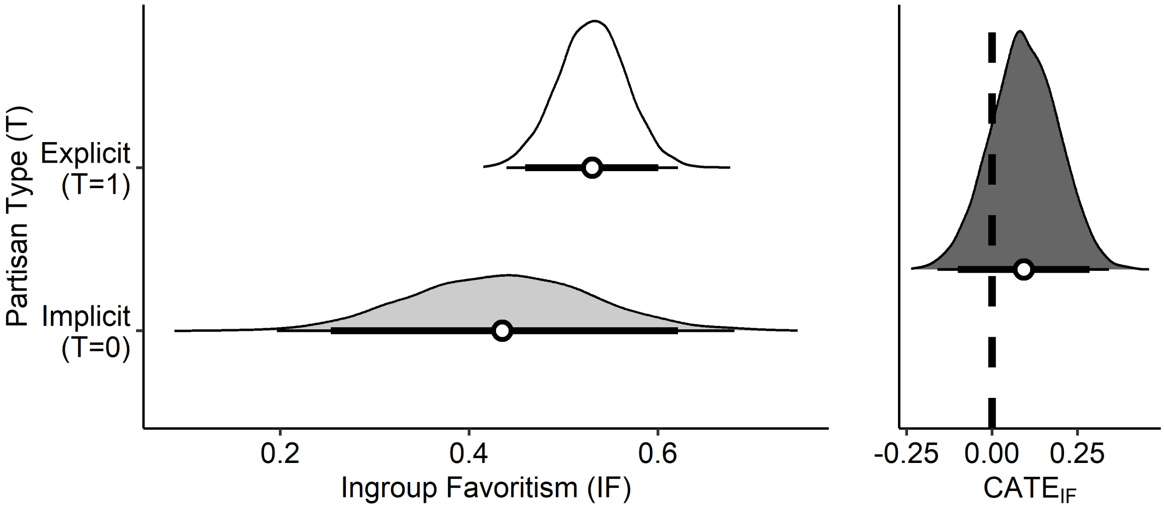

Figure 4: A title

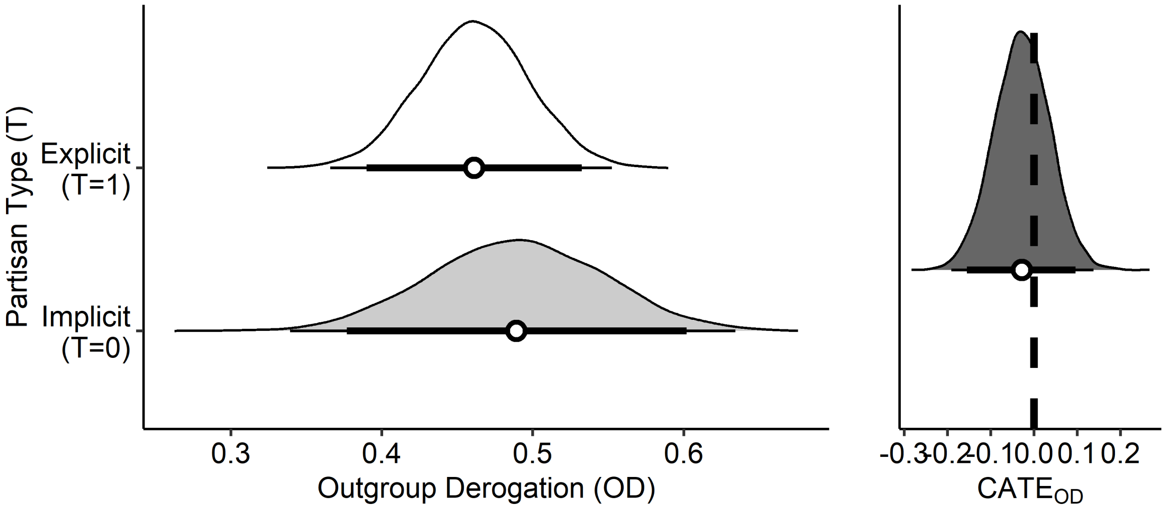

Figure 5: A title

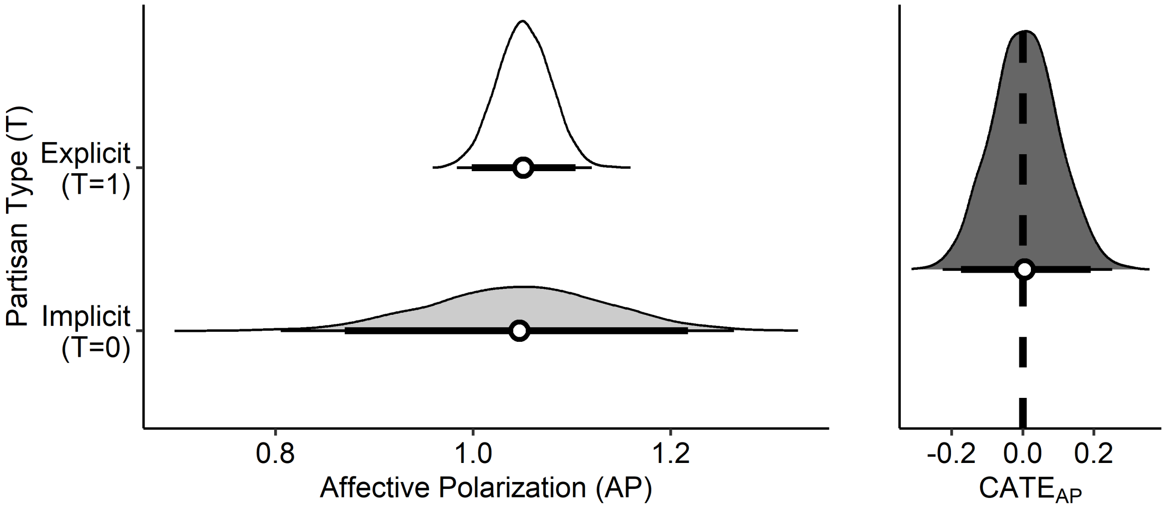

Figure 6: A title

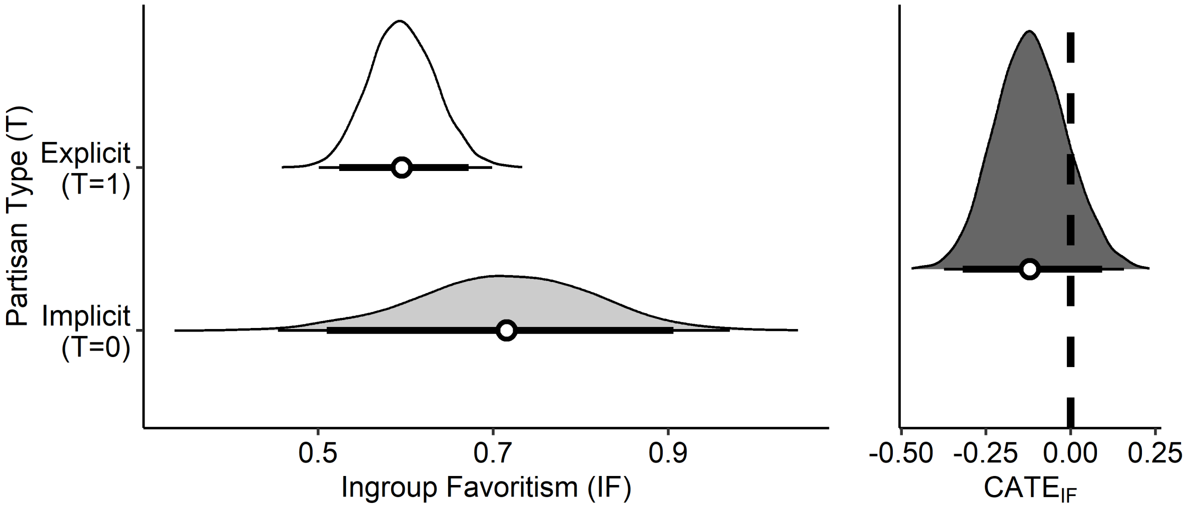

Figure 7: A title

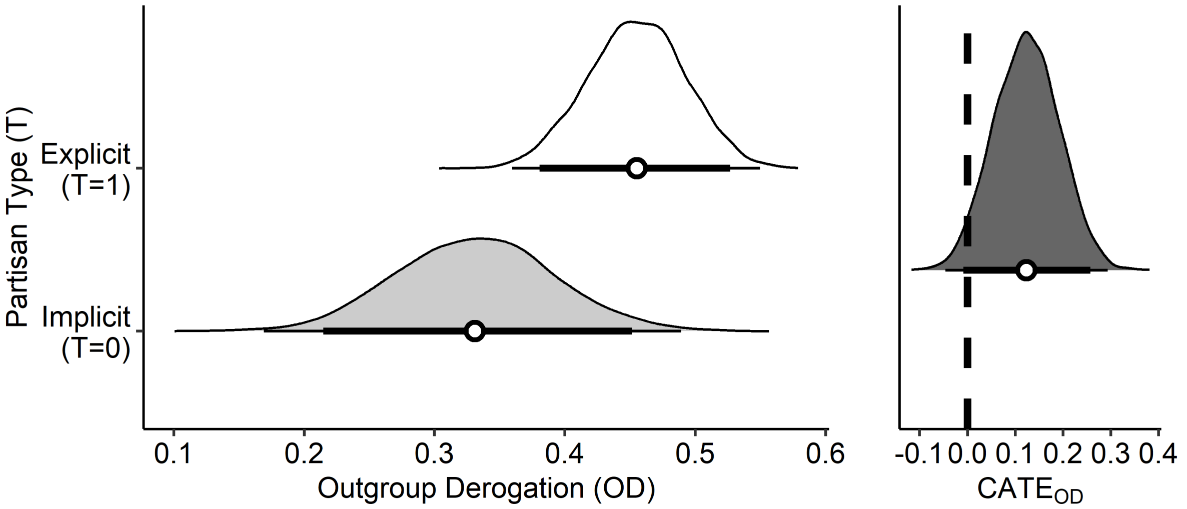

Figure 8: A title

4.2 Country analysis

5 Robustness

other response family? simple ols without any hierarchical structure

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. https://doi.org/10.1017/S0003055423000266.

Iyengar, Shanto, Gaurav Sood, and Yphtach Lelkes. 2012. "Affect, Not Ideology: A Social Identity Perspective on Polarization." *Public Opinion Quarterly* 76 (3): 405–31. https://doi.org/10.1093/poq/nfs038.

Leeper, Thomas J, Sara B Hobolt, and James Tilley. 2020. "Measuring Subgroup Preferences in Conjoint Experiments." *Political Analysis* 28 (2): 207–21. https://doi.org/10.1017/pan.2019.30.

8 Appendix