Ticking all the boxes?
A comparative study of social sorting and affective polarization

Eelco Harteveld

Abstract

In the United States, the growing hostility between supporters of the Democrats and Republicans is well-established. Research shows that such 'affective polarization' exists in other countries too, but to varying degrees. What makes compatriots with different political preferences opponents in some contexts and enemies in others? This manuscript shows, using two complementary studies, that a contributing factor is *the alignment of political and non-political identities*. First, I employ CSES data to predict the level of affective polarization by such 'social sorting' at 119 elections in 40 countries, showing that greater overlap of partisan divisions with non-political divisions in a society (along lines of income, education, religion and region) is associated with more dislike of political outgroups. Second, using Dutch panel data I show that individuals who fit the socio-demographic 'profile' of their party better tend to be more affectively polarized. This has important implications for our understanding of affective polarization.

Keywords: affective polarization; social sorting; political identities; political behavior

Funding: This work was supported by the *Dutch National Science Council* (NWO)

Introduction

Politics seems increasingly tribal and divisive. This has spurred scholarly interest in the phenomenon of *affective polarization*, or antipathy between compatriots with opposing political identities (Iyengar et al. 2012; Iyengar & Westwood 2015; Mason 2015; Mason 2018; Abramowitz & Webster 2016; Iyengar et al. 2018; Wagner 2017; Banda and Cluverius 2018; Reiljan 2019). This mutual resentment was (infamously) visible in the campaigns and aftermath of the Trump and Brexit victories (Hobolt et al. 2018; Abramowitz & McCoy 2019). A number of recent studies (Iyengar and Westwood 2015; Wagner 2019; Gidron et al. 2019a; Reiljan 2019; Lauka 2018) have demonstrated that affective polarization exist outside the US and UK context too – sometimes even more viscerally so. Affective polarization can erode citizens' willingness to engage with opposing political views, to accept others' democratic claims, and ultimately even to accept defeat in elections (Strickler 2017; Hetherington & Rudolph 2015; Tappin & McKay 2019). Taken to its extremes, affective polarization can spur dehumanization (Martherus et al. 2019) and lower the bar for political violence (Kalmoe & Mason 2018).

While present across the globe, the extent to which societies are affectively polarized varies widely between countries and over time (Iyengar and Westwood 2015; Reiljan 2019; Boxell 2020). This shows that mutual dislike across political camps is not just a 'sign of the times', but rather something that thrives under certain conditions but not others. Gidron et al. (2019a) therefore call for more *comparative* research on affective polarization. This study aims to contribute to this effort by investigating a factor that might explain (macro- and micro-level) variation in affective polarization: the *alignment of political and non-political divisions in society* ('social sorting').

It is a classic insight that cross-cutting cleavages decrease social tensions (Lipset 1960). Mason (2015; 1016; 2018) demonstrates that Americans with aligned religious, racial and partisan identities – for instance, Christian and non-White Republicans – are more antagonistic towards partisan outgroups. The contribution of the present study is to test this relation on a broader set of cases. The hypothesis is that if politically like-minded individuals in any society *also* share other non-political identities – i.e., if they are *socially sorted along political lines* – they will be less tolerant towards those with opposing views. The most relevant non-political division could be between religious groups in one country; between city dwellers and rural inhabitants in another; between the poorest and the middle class in yet another. Regardless, in each of these cases,

antagonism between political camps can be expected to be amplified if they align more strongly with one or more societal fault lines. Political 'ingroups' and 'outgroups' will appear even more distinct and homogeneous, fostering negative outgroup affect and behavior (Brewer and Roccas 2002).

I investigate the relation between social sorting and affective polarization comparatively using two complementary studies, one on the aggregate and one on the individual level. First, building on measures of affective polarization created by Reiljan (2019), I analyze CSES data on 40 countries between 1996 and 2018 to investigate the relation between the alignment of political and non-political identities (on the one hand) and affective polarization (on the other). I show that if political divisions correlate more strongly with non-political ones, affective polarization tends to be higher. I supplement this with a study of individual-level panel data from the Netherlands (2007-2018; N=8205) that allows me to track the extent to which a respondent fits the socio-demographic profile of a party. I show that individuals who are more strongly 'sorted' in this way are also more affectively polarized. The Conclusion section discusses the implications for our understanding of affective polarization.

Theory

Partisanship as a (divisive) social identity

Our views on political issues – whether it is a partisan identity such as being a Social Democratic supporter, an ideological identity such as being conservative, or even an issue identity such as being a 'Remainer' (Hobolt et al. 2019) – often constitute a *social identity*. A social identity is "that part of an individual's self-concept which derives from his knowledge of his membership of a group [...] together with the value and emotional significance attached to the membership" (Tajfel 1979). Depending on intergroup dynamics, this can also involve a smaller or greater extent of negative bias towards the out-group (Brewer 1999). The resulting 'affect gap' towards the political outgroup is now commonly called *affective polarization* (Iyengar et al. 2012).[1]

[1] The word 'polarization' has an ambiguous connotation of both a *level* (i.e. a state of division) and a *process* (the increase of this division). I use the word in the former sense. Polarization can thus be increasing, decreasing, or stable.

Research in the United States has a longstanding tradition of conceiving political identities above all along partisanship lines (Campbell et al. 1960). Iyengar & Westwood (2015) therefore define affective polarization in the American context as "the tendency of people identifying as Republicans or Democrats to view opposing partisans negatively and copartisans positively". They demonstrate that this 'affect gap' has grown steadily. 'Feeling thermometers' show that Americans feel increasingly *colder* towards adherents of the opposite party. They are also increasingly unwilling to have social interactions with outpartisans – most vividly illustrated by the steep increase in the percentage of Americans that would dislike their son or daughter marrying an outpartisan, which increased from 4-5% in 1960 to 33% in 2010 (Iyengar et al. 2012) and to 45% in 2019 (Jones & Naijle 2019). These patterns are more than mere survey artefacts: implicit association tests (IATs) show that unconscious prejudice based on partisanship is (now) stronger in the US than negative affect based on race (Iyengar & Westwood 2015).

There are reasons to fear that sustained or growing affective polarization has a negative impact on the health of democracies. Next to competitive elections, a well-functioning democracy also requires norms that allow for deliberation and compromise between citizens and elites of opposing political camps (Ziblatt & Levitsky 2018; Strickler 2017). If political identities become strongly entrenched, and those voting for the 'wrong party' become loathed, such acceptance (or at least pragmatism) might dwindle. Indeed, a recent study suggests that a nonnegliable share of Americans would see violence legitimated if the opposing party wins in 2020 (Kalmoe and Mason 2018).[2]

In the United States, the object of most research so far, the nature of affective polarization is shaped by the dynamics of its two-party system, which in turn increasingly overlaps with ideological divisions (Baldassarri & Gelman 2008; Webster and Abramowitz 2017; Banda and Cluverius 2018). In many other countries, the ballot offers numerous parties, and self-identification with any single party is generally weaker (though not absent; Huddy et al. 2018; Bankert et al. 2017). In those contexts, a partisan ingroup is not opposed to every possible partisan outgroup to the same extent (Wagner 2019). 'Political identities' in the broad sense of the word also reflect ideological identities (Devine 2015; Malka & Lelkes 2010) and even issue identities (Hobolt et al. 2019). Mason (2018: 878) shows that Americans, too, "are dividing themselves

[2] Although, on the other hand, it might also have positive consequences for political engagement, including higher turnout (Wagner 2019) and correct voting (Pierce and Lau 2019).

socially on the basis of whether they call themselves liberal or conservative, independent of their actual policy differences". A supposedly shared ideological identification as 'Left-wing' might lead a Social Democratic supporter to feel more warmly towards supporters of Green parties than towards supporters of a mainstream, let alone radical, right-wing party. As Reiljan (2019) and Wagner (2019) show, this provides not only an ingroup but also an outgroup, which tend to disliked.

It is crucial to recognize that affective polarization is not the same as ideological polarization – the extent to which citizens disagree on matters of ideology and issue positions. This relation is endogenous and not straightforward (Iyengar et al. 2018; Ward and Tavits 2019). While affective polarization has grown according to most accounts, there is less evidence for surging ideological polarization: regarding most topics, Americans' and Europeans' actual views have become less, rather than more, divided (Baldassarri & Gelman 2008; Adams et al. 2012; Nuesser et al. 2014). Indeed, Reiljan (2019) and Gidron et al. (2019a) find a weak or inconsistent relation between ideological polarization on affective polarization. Individual-level evidence is mixed too (Rogowski & Sutherland 2016; Bougher 2017). All in all, it is clear that affective polarization depends partly – perhaps mostly – on other factors than the strength of ideological disagreement between camps.

The role of social sorting along political lines

This paper builds on the work by Mason (2015; 2016; 2018) by studying the role of *social sorting along political lines* ('social sorting' in short[3]), or the alignment of political identities with non-political identities. The general expectation is that such alignment fosters antipathy towards political opponents. First of all, it infuses political divisions with the tensions characterizing other divisions in society. If a political outgroup (say, Green party supporters) is consciously or subconsciously associated with non-political outgroups (say, the highly educated or city-dwellers), this amplifies any purely *political* antipathy (say, over environmental policy) with tensions that might exist between the higher and lower educated, or between urbanites and rural inhabitants, based on allegedly competing values, lifestyle or interests.

[3] The term 'sorting' has been used to denote a range of related phenomena, most importantly (in the US context) the overlap of ideology and party identities. This is of interest in the US case, where such overlap has been historically low until recent decades, but less so in the context of this comparative paper. Rather, I am interested in the overlap of political identities with non-political identities.

Secondly, from a Social Identity Theory (SIT) perspective, social sorting makes political outgroups more easily identifiable and stereotypical (Roccas and Brewer 2002). Imagine a situation in which a person associates a political outgroup (say, again, Green party supporters) to typically be an outgroup in many other respects as well (say, university graduates, atheists and city dwellers). In that case her or his 'social identity complexity' can be said to be *low*. Reversely, if the same person would perceive that the political outgroup (Green party supporters) is often on the ingroup side in other respects ('us lower educated' or 'us rural folk'), this would signal a *higher* 'social identity complexity'. A more complex social identity generally enhances tolerances for outgroups (Roccas and Brewer 2002). It decreases *intracategory assimilation* and *intercategory accentuation* – in other words, the tendency to overestimate the internal homogeneity of, as well as differences between, the ingroup and the outgroup.[4] As a result, groups such as 'Green voters' and 'Conservative voters' will appear less like a threatening homogeneous and distinct block. Indeed, stressing (in experimental settings) that outgroup members are ingroup members on another dimension decreases bias (Gaertner et al. 1993; Levendusky 2018). It follows from this that having cross-cutting political and non-political identities should reduce negative affect towards the political outgroup.

In the US context, Mason (2015; 2016; 2018) demonstrates that social sorting is indeed associated with stronger affective polarization. She shows that Republicans who are also White and Christian are more affectively polarized than Republicans with a cross-cutting identity – regardless of ideological extremity. The same mechanism is observable among Democrats who identify as secular and non-White, as opposed to Democrats with cross-cutting identities. Mason and Wronski (2018) therefore stress the need to look the "cumulative relationship between social identities and partisan identities". The fact that affective polarization depends on the *configuration* of a broader set of identities is also apparent from the work of Levendusky (2018). He finds affective polarization to be lower among respondents who were cued to be aware of their overarching *American* identity and among respondents interviewed on the national holiday July 4th (which signals that the *political outgroup* is a at the same time a *national ingroup*; but see Brandt and Turner-Zwinkels 2020).

[4] Furthermore, disliking an outgroup that is simultaneously often perceived as an ingroup would create cognitive inconsistencies that need to be avoided (Roccas and Brewer 2002)

Of course, the fact that non-political attributes tend to generally go together with a particular political position does not mean that individuals will universally perceive it as such. Roccas and Brewer (2002) and Mason (2018) therefore also study the cognitive representation of such alignment.[5] Still, research in US shows partisans are generally aware of, and often even grossly overestimate, the features associated with Democratic or Republic support, such as religion, union membership or income (Ahler and Sood 2018). In short, the alignment of social groups with political camps (the empirical focus of this study) can be expected to feature systematically, if imperfectly, in citizens' perceptions. Through the mechanisms discussed above, this is expected to foster affective polarization.

> Hypothesis: *higher levels of social sorting along political lines are associated with higher levels of affective polarization*

The extent to which political groups are socially sorted will vary between contexts and over time. Spatial segregation can create neighborhoods or entire regions in which citizens share not only a worldview but also other social identities such as education or ethnicity (Bishop 2008; but see Abrams and Fiorina 2012). Cross-cutting institutions that socially integrate socio-demographic groups and a plethora of political views – for instance churches or sport clubs – can decline (but also grow) in importance or become organized along more homogeneous (or heterogeneous) lines. In turn, affective polarization might stimulate people to 'sort', i.e. withdraw in homogeneous groups, which would make the relation endogenous. Establishing the causal direction with certainty is beyond the observational data presented in this paper. While the US studies discussed above provided strong evidence for the hypothesized mechanism, any observed correlation between affective polarization on social sorting might partly reflect a reciprocal effect. Nevertheless, it remains important to study whether the relation between social sorting and affective polarization travels beyond the US context.

[5] For instance, Roccas and Brewer (2002) show that under conditions of stress people's social identities become less complex across the board.

Design

In the remainder of this paper, I investigate the relation between social sorting and affective polarization empirically in two complementary studies. The first study does so on the *aggregate* level using the *Comparative Study of Electoral Systems* (CSES) data, covering 119 elections in 40 countries between 1996 and 2018. I calculate a *Social Sorting Score* inspired by Selway (2011) that captures the extent to which political divisions align with non-political cleavages in a society. I then use the *Social Sorting Score* to predict affective polarization, which I measure following the operationalization of Reiljan (2019). The time span provides sufficient variation *within countries*, which isolates the correlations as far as possible from (relatively) time-invariant confounding factors such as political system or culture.

The repeated cross-sectional nature of the CSES data makes it less suitable to make strong inferences about correlations on the individual level. I therefore complement my design with the second study, which employs panel data to study the mechanism on the individual level. I use the population-representative Dutch LISS data that allows to track the social sorting of individuals over a period of up to ten years (2008-2018). I develop a measure of the extent to which individuals fit the socio-demographic profile of a party, and use a fixed effects panel regression to study if affective polarization moves in tandem with an individual's social sorting.

Study 1: affective polarization and social sorting in 40 countries

Data and cases

The *Comparative Study of Electoral Systems*[6] data consists of harmonized election studies in countries around the world between 1996 and 2018. This provides a unique collection of equivalent survey items across all continents over 20 years, collected at a moment when political identities are most salient and impactful – elections. 40 countries have valid observations of the key independent

[6] www.cses.org

and dependent variables at least once (see Appendix A in the Supporting Information).[7] In total this yields a maximum of 119 (country-election) observations.

Operationalization

Affective polarization. The dependent variable is Reiljan's (2019) *Affective Polarization Index* (API). Like the measures proposed by Wagner (2019) and Gidron et al. (2019a), this is based on sympathy scores towards out-parties. Although related, this is not the same as affect towards *partisans*. However, because of the correlation between the two (r = 0.69 in the US in Iyengar et al. 2012; see also Druckman et al. 2019) it can be used – with some caveats – to study affective polarization. The measure also reflects ideological polarization (Wagner 2019), so I control for ideological polarization in each model (see below).

The *Affective Polarization Index* can be thought of as a 'weighted sympathy' towards outparties. The formula below (taken from Reiljan 2019: 5-6) summarizes the procedure. First, for each partisan group n (i.e. respondents who say they identify with a particular party[8]) the average evaluation towards all other parties m is subtracted from their evaluation of their inparty. Each difference is weighted by the relative size of a party (measured as vote share) and summed over all outparties. This is repeated for each of the partisan groups, and all these 'relative AP' scores are weighted by vote shared and summed up as well.

$$\textit{Affective Polarization Index}_c = \sum_{n=1}^{N}\left[\sum_{\substack{m=1 \\ m \neq n}}^{N}\left((\text{Like}_n - \text{Like}_m) \times \left(\frac{\text{Vote share}_m}{1 - \text{Vote share}_n}\right)\right) \times \text{Vote share}_n\right] \quad (1)$$

Social sorting. The operationalization of the main independent variable is inspired by the work on cross-cuttingness by Selway (2011). Cross-cuttingness is the reverse of sorting, and can be defined as the extent to which "group i on cleavage x is identically distributed among groups on cleavage y with all other groups on cleavage x" (Selway 2011: 51). Selway (2011) measures the cross-

[7] I opted not to restrict the analysis to a particular subset, because a priori the mechanisms should apply in all contexts. An important caveat is the inclusion of 'democratic authoritarian' societies. Excluding countries denoted as 'Not free' or 'Partly free' according to Freedom House in one or more of the elections (Russia, Turkey, Serbia, Ukraine, and Hungary) does not affect the results substantively.

[8] Based on the question "Do you feel closer to any particular party?". A replication based on vote choice yields very similar results.

cuttingness of two divisions by calculating Cramér's *V*, a measure of association for categorical variables, and subtracting it from 1. Because I am interested in sorting, I simply use Cramér's *V*.[9]

For each of the elections covered in the dataset, I calculated four Cramér's *V* associations between on the one hand *party identification*[10] and on the other hand each of the following non-political variables: *income* (5 quintiles), *education* (5 harmonized categories), *region* (number of categories differs per country), and *religion* (including 'not religious'; number of categories differs per country).[11] The selection of non-political divisions was restricted to variables available in almost all election studies.[12] Still, the above set captures major social fault lines that are often expressed politically: class (approximated by income), the 'new class' or 'winners and losers of globalization' division (education), as well as center-periphery (region) and religious oppositions. Each of the four Cramer's V scores reflects how well party identification can be predicted by the respective non-political variable. I calculate an overall *Social Sorting Score* for a context *c* as the average of the associations across the four dyads:[13]

$$\textit{Social Sorting Score}_c = \frac{\begin{array}{r}(\textit{Cramer's } V_{\textit{ideology, income}} + \\ \textit{Cramer's } V_{\textit{ideology, education}} + \\ \textit{Cramer's } V_{\textit{ideology, region}} + \\ \textit{Cramer's } V_{\textit{ideology, religion}})\end{array}}{4} \qquad (2)$$

This measure is a priori agnostic about the *shape* of the relation between party preference and each of the other identities. For instance, in some countries, the lower educated are relative often supporters of the Social Democratic party, in other countries of the Conservative party; in again other countries, they might usually vote for far left or far right alternatives, rather than centrist ones. All of these patterns are captured by the cross-cuttingness measure: it reflects how well

---

[9] Cramer's *V* reflects the divergence in a crosstab between, on the one hand, the *expected* number of observations in each cell (say, a vote for 'Greens' and 'higher educated') if there were to be *no* association, and on the other hand, the actually *observed* observations in a cell. The larger this (absolute) difference, the better the one variable predicts the other.

[10] Also based on the question "Do you feel closer to any particular party?". Here too, a replication with vote choice yields very similar results.

[11] Only parties for which at least 25 respondents voted were included. As an alternative specification, the analysis was replicated using ideological self-placement rather than party preference; this yielded very similar results.

[12] For this reason, ethnicity could not be included.

[13] I opted to take the mean score, rather than for instance the maximum among any of the four, because a cumulation of *multiple* alignments is likely to be especially polarizing (Mason 2015). In case one of the variables is missing in particular country or election, the *Social Sorting Score* is calculated over the remaining variables. This maximizes the number of cases, although it might add further cross-national variation in the score. However, this is less problematic because patterns of missingness are usually country- rather than election-specific, and the focus of the analysis is on within-country effects.

categories of education (or income, region, or religion) predict vote choice. The better it does, the higher the *Social Sorting Score*.

To be sure, this measure relies on ascribed identities that can be 'objectively' measured (education, income, region, and religion), rather than directly assessing citizens subjective identities. However, it seems plausible that the factors included in the measure have some systematic bearing on many citizens' subjective identities. Still, the *Social Sorting Score* will be sensitive to the extent to which the chosen variables capture the relevant divisions in a country. Where it doesn't, a country might still be sorted along other lines, in which case the measure will *underestimate* the actual level of sorting. For this reason, within-country comparisons will be most valid.

Control variables. As control variables, I include several other characteristics that can be expected to matter for affective polarization. Most importantly, I include *Ideological polarization on the elite level* (see Banda and Cluverius 2018) and *Ideological polarization on the mass level*. By controlling for the latter, I aim to isolate as far as possible the 'non-ideological' component of the *Affective Polarization Index*. *Ideological polarization on the elite level* is measured by the standard deviation in left-right positions according to the Comparative Manifesto Project's database (MARPOR).[14] *Ideological polarization on the mass level* is simply the standard deviation in left-right positions of respondents in a given year and country.

Because my main interest is a comparison within countries over time, I chose not to include slower moving economic indicators such as GDP or inequality (Gidron et al. 2019a). I do include the control variable *Effective number of parties* as developed by Laakso and Taagepera (1979).[15] A larger set of parties changes the logic of competition and cooperation, and possibly reflects a more consensual political culture. Arguably, this indicator too varies more substantially between countries than over time. Still, I include it because I expect it to correlate with the outcome variable in a mechanical way, too: the more parties are available in the questionnaire, the more negative affect towards one of them gets averaged out by others. This indicator should absorb much of that correlation.

[14] See manifesto-project.wzb.eu.

[15] Based on electoral (rather than parliamentary) strength.

Furthermore, I include the *Salience of cultural issues* and *Salience of economic issues*. Gidron et al. (2019b) show that distance on the cultural dimension is especially conducive to affective polarization. My measure is based on manifesto data collected by MARPOR, and consists of the share of the manifesto's of all parties in a given election that is devoted to cultural issues or economic issues, respectively.[16]

Design

In total, the independent and dependent variables are available for 119 observations in 40 countries. Because both affective polarization and social sorting are measured at the level of country-election, I will perform the analysis at the aggregated level, using both a fixed effects (FE) and between-effects (BE) estimation in a panel regression. The FE specification restricts the estimation to variation *within countries, between elections*. In other words: at elections in which a country is more sorted than usual, is affective polarization relatively high too? This absorbs potential time-invariant confounders at the country level, and also reduces the problem that the values on *Social Sorting* are likely to be influenced to a substantive extent by whether the chosen nonpolitical divisions matter in a particular country. The between-effects (BE) specification is based on variation *between countries*. Do those countries that are (on average) more sorted also experience more affective polarization? Of the 40 countries under study, 10 appeared in the dataset only once; these are only included in the between-effects analyses.[17]

Results

Descriptives. Appendix B reports the trends in affective polarization, social sorting, and ideological polarization on the mass level. It testifies that affective polarization has different levels and trends across the world. Replicating Reiljan (2019), Southern Europe and East-Central Europe generally show very high scores, followed by the US. Furthermore, there is ample variation in the trend of social sorting (see Boxell et al. 2019). It also shows that affective polarization and ideological polarization tend to move often, though far from always, in tandem. The same is true for affective

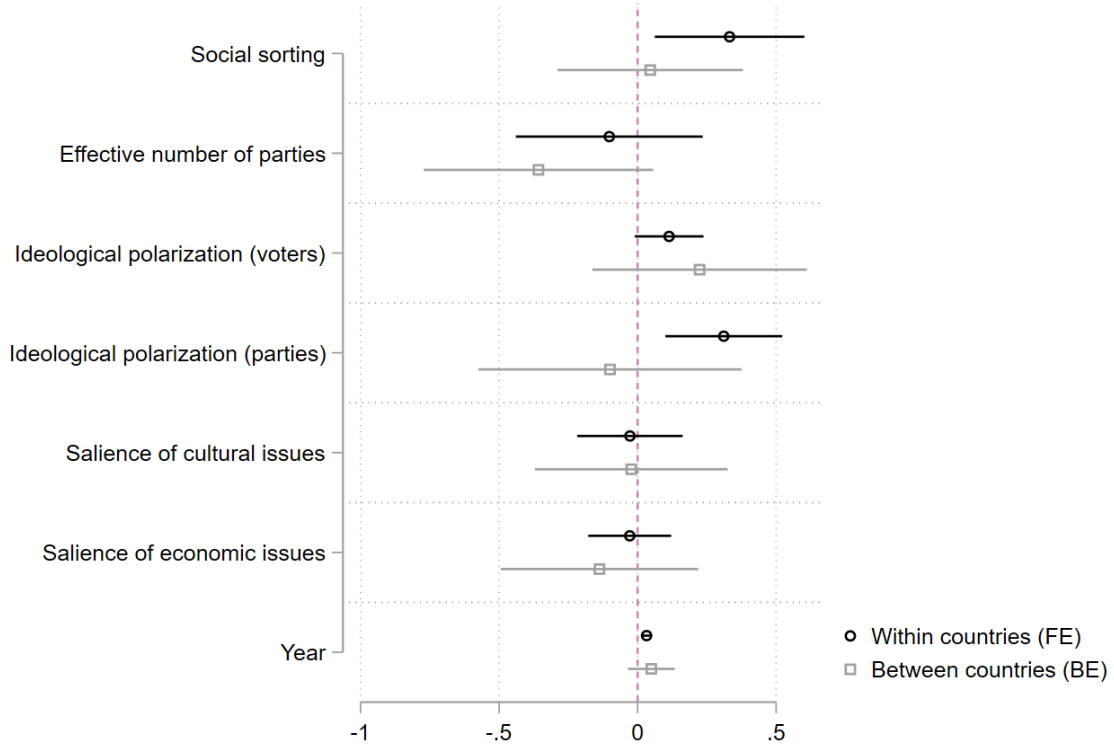
[16] Coded as cultural issues were: environmentalism, culture, equality, national way of live, law and order, multiculturalism. Coded as economic issues were: free market economy; incentives; market regulation; economic planning; corporatism; protectionism; economic goals; demand management; economic growth; controlled economy; nationalization; economic orthodoxy; Marxist analysis.

[17] The average number of observations for a country is 2975.

polarization and social sorting. However, the large number of cases defies patterns to be easily observed visually. For that we turn to a regression analysis.

Regression. Figure 1 shows regression coefficients of two regression models (a fixed-effects and a between-effects model). To allow for some comparability, both the *Affective Polarization Index* and the independent variable were standardized across the sample as a whole. Each model contains all control variables as well as the variable *Year* to capture possible trends in the dependent variable. Table 1 in Appendix C shows the full regression tables.

FIGURE 1. REGRESSION COEFFICIENTS OF FE AND BE MODELS EXPLAINING *AFFECTIVE POLARIZATION INDEX*



$N_{observations}$ = 119; N_{groups} = 40
Note: 95% confidence intervals
Source: CSES

The black dot (which denotes the effects of a within-country FE model) shows that the *Social Sorting Score* has a positive effect (p = 0.025) on the *Affective Polarization Index*. The effect size is substantial: an increase of 1 standard deviation in social sorting is associated with a 0.33 standard deviation increase in affective polarization. The grey, between-country coefficient, by contrast,

provides no evidence for an effect. In other words, variation in social sorting within countries predicts variation in affective polarization in the same country; but this is not the case between countries. The latter might reflect that scores on the social sorting measure do not compare well between contexts, given that the selection of non-political divisions it includes is not equally relevant everywhere. It could also mean, plausibly, that many unobserved factors determine the level of affective polarization in a country that add noise to the correlation. At any rate, the existence of a within-country effect supports the hypothesis that social sorting goes together with higher affective polarization outside the US, too.

The other variables perform largely as expected, though not all of them are robust predictors. A larger effective number of parties is associated with lower affective polarization, but only (and only significantly so at the 10% level) between countries than within countries (which is not surprising for a slow-moving variable). To the extent that we can draw inference from the effect, it appears that the more parties there are in a system, the lower affective polarization is. As mentioned in the operationalization section, this is theoretically plausible but might partly be a methodological artefact.

Ideological polarization, both on the level of voters and parties, is associated with more affective polarization. Ideological polarization among voters is associated with more affective polarization when comparing both between and (especially) within countries; ideological polarization among parties only explains variation within countries. It is notable that the *Social Sorting Score* has an effect size that is comparable to that of ideological polarization. This confirms that affective polarization is rooted in much more than substantive disagreement alone (Iyengar et al. 2018). The salience of either cultural or economic issues does not play a role, in contrast to Gidron et al. (2019b). The significant positive of *Year* tells us that a small upward trend in affective polarization exists, even net of the included predictors.

I conducted additional analyses to gauge the robustness of the main findings (reported in Tables 2 and 3 of Appendix C). First, I estimated a first-differences model (i.e. testing whether the *change* in API is predicted by the *change* in the independent variables), which, too, isolates the effect from time-invariant confounders but is more robust in case of high serial correlation. This yielded a significant effect of similar magnitude ($b = 0.35$; $p = 0.023$). Second, to test the sensitivity of the findings for the patterns in individual countries, I replicated the fixed-effects analysis using a 'jackknife' strategy, removing one country at a time. Under this more conservative

specification the *p*-value of the effect of social sorting increases from 0.025 to 0.062. Given the directional hypothesis I continue to interpret this as evidence for the hypothesis.

Study 2: a Dutch panel study on affective polarization and sorting

Data and case

In Study 1, we learned that if societies are more sorted along political lines, affective polarization is generally higher. In this study, I zoom in on one case, the Netherlands, and track the 'social sortedness' and affective polarization among a total of 8205 unique individuals over periods up to 10 years (from 2008 to 2018). If the pattern of Study 1 is brought about by the theoretical mechanism discussed earlier, we would expect to see that individuals who fit their ingroup party better in socio-demographic terms should generally be more affectively polarized.

The Netherlands experiences only weak affective polarization according to Reiljan (2019) and is not characterized by the unique linguistic, ethnic or religious cleavages studied by Westwood et al. (2015). This thus makes for a less-likely case to explain any variation in affective polarization by social sorting. To the extent that the two do correlate, it is plausible that effects are at least as pronounced in contexts that are both more affectively polarized and sorted.

The source of the data is the *Longitudinal Internet Studies for the Social sciences* (LISS) panel, which is recruited based on a population-representative sample.[18] Its respondents answer a battery of questions on politics yearly. In total, 8205 unique individuals with nonmissing observations took part during a part or the whole of the period 2008-2018. 69% of the respondents has non-missing responses to the relevant questions for at least 3 waves; 59% for at least 4 waves; and 48% for 5 waves or more (on average 3.4 waves). See Appendix D for descriptive statistics of the variables.

Operationalization

Affective polarization. Because, in contrast to Study 1, the unit of analysis is the individual, Reiljan's (2019) macro-level *Affective Polarization Index* is not a feasible indicator. Instead, affective

[18] The panel is based on a true probability sample of households drawn from the population register. Households that could not otherwise participate are provided with a computer and Internet connection. See www.lissdata.nl for more information.

polarization *of an individual* is measured using Wagner's (2019) *Weighted Affective Polarization* measure. Like Reiljan's (2019) measure, this is a 'weighted sympathy' measure, but calculated based on each *individual's* (rather than each *partisan group's*) sympathy scores towards all parties. This is achieved by taking a respondent's sympathy towards each party, subtracting it from the (weighted) average sympathy towards all parties, taking the squared term of this (to make it absolute), and multiplying it by the size of the party $v_{p.}$ The final score is the square root of the sum of this score across all parties. The full equation is[19]

$$\textit{Affective Polariation}_i = \sqrt{\Sigma_{p=1}^{P} v_p(like_{ip} - \overline{like_i})^2} \quad (3)$$

The element $\overline{like_i}$ is the weighted sympathy towards all parties:

$$\overline{like_i} = \Sigma_{p=1}^{P}(v_p * like_{ip}) \quad (4)$$

Social sorting. In the two-party context of the United States, Mason (2015) measured sorting by scoring, for a set of identities such as race and religion, whether each respondent is 'sorted' on it or not (i.e. belongs to the modal group of the respondent's in-party), weighted by the strength of the identification. The strength of the actual identification is not available in the LISS data, so this study – like Study 1 – relies on 'objective', rather than subjective, identities. Moreover, in a fragmented multiparty context, it is less straightforward to identify whether an individual is sorted or not because not all identities are equally relevant in predicting each party.[20] Ideally, a measure is therefore agnostic about the factors on which sorting takes place.

I therefore calculate social sorting as *the extent to which an individual's party preference can be successfully predicted by socio-demographic variables*. I first estimate, for each wave and each party, a model in which I predict the vote by age, education, income, and urbanity.[21] I then

[19] Copied from Wagner (2019).

[20] A Christian democratic party might be sorted above all by religion, a social democratic party by class, a green party by education, etcetera.

[21] In contrast to Study 1, this variable includes urbanity rather than region, because urban-rural differences in political support are generally larger than regional differences in the small and dense country of the Netherlands. Religion is not available in all waves and therefore excluded. Age is used to capture the age differential in support for parties that was visible in recent elections (Van der Meer et al. 2019)

estimate, for each individual, their *residual* according to this model.[22] This larger the *absolute* residual, the worse the individual 'fits' the socio-demographic composition of a party (as her or his support is either under- or overestimated). The *Social sorting* score, which should reflect a *good* fit to a party electorate in socio-demographic terms, is one minus the average absolute residual.

Control variables. Control variables include dummies of left-right positions (to capture ideological polarization on the individual level; recoded to Far left [0-2]; Moderate left [3-4]; Center [5]; Moderate left [5-6]; and Far right [7-10]) and political interest (three categories). I also control for education, age and gender. I formulate no expectations for the latter variables, but for their role in the US see Iyengar et al. (2012: 425). These latter variables do not (or barely) change over time and are thus included in the between-respondent model only.

Design

As in Study 1, the goal of the analysis is to establish the relationship between social sorting and affective polarization. I focus the main analysis on explaining both variation *between* and *within* respondents, as the latter makes most use of panel data by isolating effects from confounding by time-invariant individual characteristics. Of course, this raises the question whether individuals *change* in their level of sorting, as the features on which sorting is based often do not change (although they sometimes do). However, respondents can switch parties and end up in a more (or less) sorted party, and the party they belong to can change in composition to become more or less like the respondent. Even if sorting would vary more between individuals than within individuals, investigating the latter allows to make inferences that are isolated from potential time-invariant individual confounders.

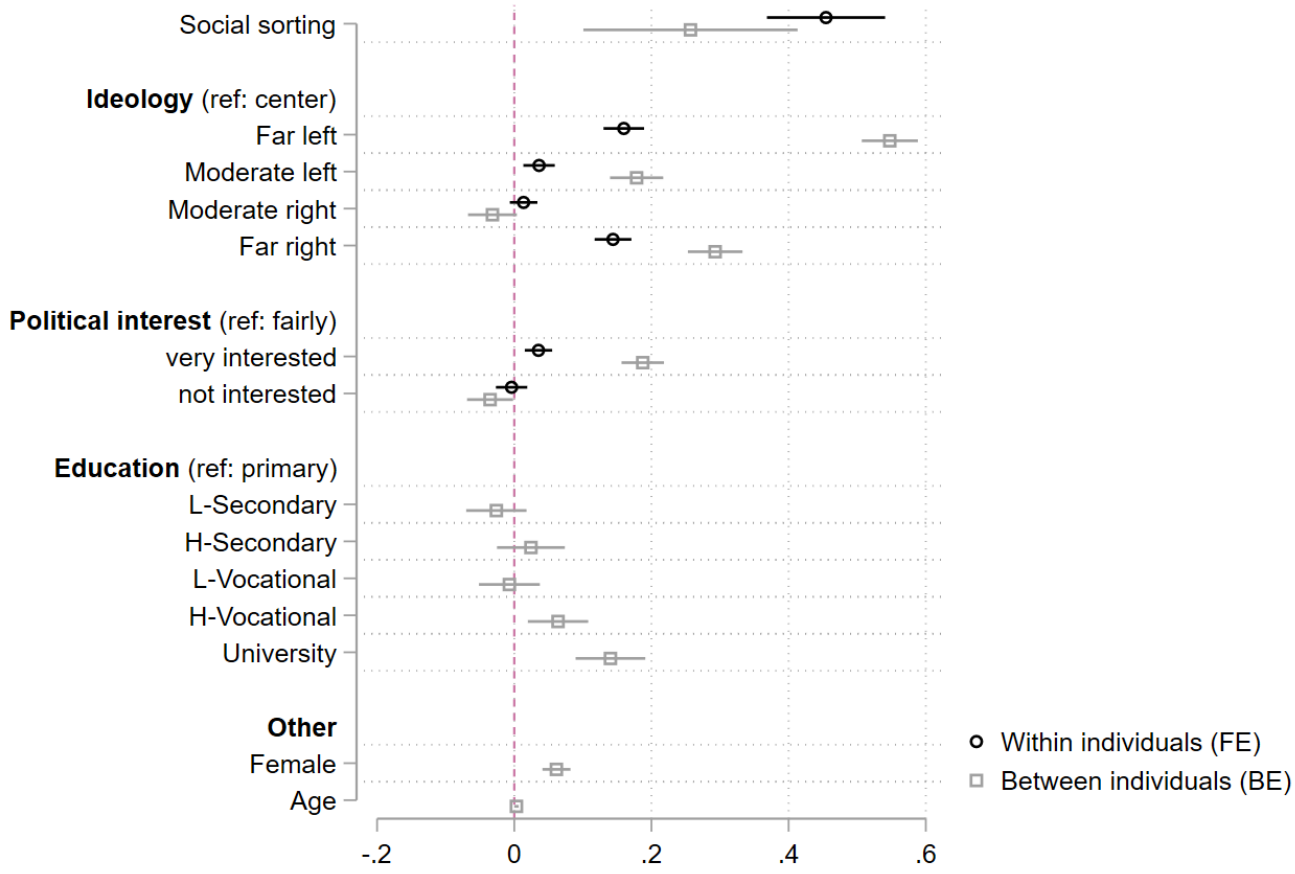
Results

Regression. Figure 2 shows the unstandardized coefficients[23] of two panel regression models (fixed effects and between effects) predicting the *Weighted Affective Polarization* score by *Social sorting* and other variables. Table 1 in Appendix E provides the full regression tables.

[22] This residual is calculated based on an OLS rather than logistic model to obtain continuous residuals. The result is a roughly normally distributed absolute residual.

[23] The coefficients are not standardized because there is no comparison of effect sizes between continuous variables, in contrast to Study 1.

FIGURE 2. REGRESSION COEFFICIENTS OF FE AND BE MODELS EXPLAINING *WEIGHTED AFFECTIVE POLARIZATION*



$N_{observations}$ = 27970; $N_{respondents}$ = 8205
Note: 95% confidence intervals
Source: LISS

The black dots show the results of a fixed effects model. The first coefficient confirms that in waves in which individuals score relatively high on *Social sorting*, this person also tends to score higher on *Weighted Affective Polarization*. Or, in other words, the better an individual's support for a party can be predicted by her or his socio-demographics, the more disliking this individual generally is of political outgroups. In the between-subjects model (grey dots), the effect of *Social sorting* is in the same direction and about half the size. Figure 2 is therefore in line with the hypothesis. If respondents start to resemble their party's profile better, they are also generally more affectively polarized; and keeping a range of socio-demographics and ideology constant, those respondents who are most sorted are also generally those who are most affectively polarized. This likely underlies the aggregated pattern found in study 1.

Respondents on the extremes (or moving to the extremes) are more affectively polarized than those on the center. When directly modelled as extremity (i.e. distance from the center score of 5), the standardized effect of extremity is 0.13, compared to a standardized effect of sorting of 0.06. Political interest, too, correlates with affective polarization. Age does not significantly predict affective polarization (in contrast to the findings among Americans in Iyengar et al. 2012), whereas women score somewhat higher (in line with *idem*), as do the higher educated (not in line with *idem*).

As an alternative specification, I again ran a first difference model, regressing change on change (see Table 2 in Appendix E). This results in a significant effect of a comparable size (b = 0.57, $p < 0.001$).

Conclusions

Cross-cutting cleavages have long been argued to decrease social tensions (Lipset 1960). Building on the research of Mason (2015; 2016; 2018), this study investigated whether – in a broader set of cases than hitherto studied – citizens' dislike towards political opponents is stronger if political identities align with non-political identities. Comparing 119 elections in 40 countries over two decades, I found that social sorting along political lines is indeed associated with affective polarization across a range of specifications. The effect size is substantive and similar to that of ideological polarization – the extent to which citizens disagree substantially on issues. I furthermore show, using Dutch panel data that allows to make stronger claims on the individual level, that citizens who fit the socio-demographic profile of their party better are more affective polarized. It is telling that this mechanism is visible even in a country with very weak affective polarization (Reiljan 2019). Both studies strongly suggest that the hostility fostered by political identities depends on its interplay with non-political identities.

In line with previous work, this study thus confirms that (growing) affective polarization is about more than just (growing) issue disagreement (Iyengar et al. 2018; Reiljan 2019; Lelkes 2018). It raises a question with important implications: how purely *political* is affective polarization? To the extent that affective polarization is partly a reproduction – or amplification – of tensions smoldering outside the realm of politics, there are limits to the extent it can be abated

by reforming a political culture or adjusting political institutions. Efforts to prevent affective polarization from becoming 'toxic' (or even lethal; Kalmoe and Mason 2018) would need to grapple with other dimensions of social divisions and segregation too. Empirically, it is difficult to isolate the relative role of the political and the non-political using observational data, precisely because citizens have conscious and subconscious associations between political and non-political groups. Experimental or conjoint designs, such as employed by Helbling and Jungkunz (2019), are important steps forward in this respect. Does flagging a non-sorted identity (i.e., portraying somebody as a Green supporter and lower educated) as opposed to a sorted identity (Green supporter and higher educated) reduce affective polarization towards that person?

The present study took a relatively coarse approach by assuming respondents to identify with the 'objective' categories presented to them in the questionnaires, and are aware of how they are aligned with the political camps in the population at large. As Ahler and Sood (2018) show, real associations between political and nonpolitical groups are reflected in people's perceptions, although in a heavily distorted way. Still, the fact remains some citizens will identify more or less strongly with – say – their class, religious, or regional group than others, and this matters for the extent to which such identities influence their views of political opponents. It is therefore plausible that the correlation between 'objective' sorting and affective polarization found in the present study is an underestimation of the complete role played by social sorting. Following Mason (2015), future comparative studies on social sorting could benefit from including measures of subjective identification. Although this requires more elaborate data, it does provide for a more direct evidence of the mechanism at hand.

The *Social Sorting* measure presented in this paper suggests that sorting is (somewhat) on the rise across the globe, but with substantial variation in levels and trends between countries. This study suggests understanding the causes of social sorting along political lines is important. It also raises the question *who* tends to be sorted most. Even if only a subset of society, such as the most politically most engaged, finds itself increasingly sorted, then the associated higher levels of affective polarization might spill over to the political debate and the public at large. Next to continuing the comparative agenda, an important next step is therefore to further explore its interaction with the individual level: who is most 'at risk' of becoming affectively polarized?

Literature

Abramowitz, A. I., & Webster, S. (2016). The rise of negative partisanship and the nationalization of US elections in the 21st century. *Electoral Studies*, 41, 12-22.

Abramowitz, A., & McCoy, J. (2019). United States: Racial Resentment, Negative Partisanship, and Polarization in Trump's America. *The ANNALS of the American Academy of Political and Social Science*, 681(1), 137-156.

Abrams, S. J., & Fiorina, M. P. (2012). "The big sort" that wasn't: A skeptical reexamination. *PS: Political Science & Politics*, 45(2), 203-210.

Adams, J., Green, J., & Milazzo, C. (2012). Who moves? Elite and mass-level depolarization in Britain, 1987–2001. *Electoral Studies*, 31(4), 643-655.

Baldassarri, D., & Gelman, A. (2008). Partisans without constraint: Political polarization and trends in American public opinion. *American Journal of Sociology*, 114(2), 408-446.

Banda, K. K., & Cluverius, J. (2018). Elite polarization, party extremity, and affective polarization. *Electoral Studies*, 56, 90-101.

Bankert, A., Huddy, L., & Rosema, M. (2017). Measuring partisanship as a social identity in multi-party systems. *Political behavior*, 39(1), 103-132.

Bischof, D. 2017. New graphic schemes for Stata: plotplain and plottig. *Stata Journal* 17: 748-759.

Bishop, B. (2009). *The big sort: Why the clustering of like-minded America is tearing us apart*. Houghton Mifflin Harcourt.

Bougher, L. D. (2017). The correlates of discord: identity, issue alignment, and political hostility in polarized America. *Political Behavior*, 39(3), 731-762.

Boxell, L., Gentzkow, M., & Shapiro, J. M. (2017). Is the internet causing political polarization? Evidence from demographics (No. w23258). *National Bureau of Economic Research*.

Brandt, M., & Turner-Zwinkels, F. M. (2020). *Proximity to the July 4th Holiday Does Not Affect Affective Polarization*. Working paper available at DOI 10.31234/osf.io/7yqkd

Campbell, A., Converse, P. E., Miller, W. E., & Stokes, D. E. (1980). *The American voter*. University of Chicago Press.

Comparative Study of Electoral Systems (www.cses.org). *CSES integrated module dataset phase 2 release* [dataset and documentation]. October 17, 2019 version. doi:10.7804/cses.imd.2019-10-17.

Devine, C. J. (2015). Ideological social identity: Psychological attachment to ideological in-groups as a political phenomenon and a behavioral influence. *Political Behavior*, 37(3), 509-535.

Druckman, J. N., & Levendusky, M. S. (2019). What do we measure when we measure affective polarization? *Public Opinion Quarterly*, 83(1), 114-122.

Fiorina, M. P., & Abrams, S. J. (2008). Political polarization in the American public. *Annual Review of Political Science*, 11, 563-588.

Gaertner, S. L., Dovidio, J. F., Anastasio, P. A., Bachman, B. A., & Rust, M. C. (1993). The common ingroup identity model: Recategorization and the reduction of intergroup bias. *European review of social psychology*, 4(1), 1-26.

Gidron, N., Adams, J., & Horne, W. (2019a). Toward a Comparative Research Agenda on Affective Polarization in Mass Publics. *APSA Comparative Politics Newsletter*, 29, 30-36.

Gidron, N., Adams, J., & Horne, W. (2019b). Who Dislikes Whom? The Drivers of Affective Polarization in Multi-Party Systems. Presented a*t APSA Annual Meeting*.

Helbling, M., & Jungkunz, S. (2020). Social divides in the age of globalization. *West European Politics*, 43(6), 1187-1210.

Hetherington, M. J., & Rudolph, T. J. (2015). *Why Washington won't work: Polarization, political trust, and the governing crisis*. University of Chicago Press.

Hobolt, S. B., Leeper, T., & Tilley, J. (2018). Divided by the vote: Affective polarization in the wake of Brexit. Presented at the *American Political Science Association* General Conference, Boston.

Huddy, L., Bankert, A., & Davies, C. L. (2018) Expressive vs. Instrumental Partisanship in Multi-Party European Systems. Forthcoming in *Advances in Political Psychology*.

Iyengar, S., & Westwood, S. J. (2015). Fear and loathing across party lines: New evidence on group polarization. *American Journal of Political Science*, 59(3), 690-707.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, 22, 129-146.

Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, Not Ideology: A Social Identity Perspective on Polarization. *Public opinion quarterly*, 76(3), 405-431.

Jones, R.P and Najle, M. American Democracy in Crisis: The Fate of Pluralism in a Divided Nation. *PRRI Report* (available on www.prri.org/research/american-democracy-in-crisis-the-fate-of-pluralism-in-a-divided-nation).

Kalmoe, N., & Mason, L. (2018). Lethal mass partisanship: Prevalence, correlates, and electoral contingencies. Presented at *NCAPSA American Politics Meeting*.

Laakso, M., & Taagepera, R. (1979). "Effective" number of parties: a measure with application to West Europe. *Comparative political studies*, 12(1), 3-27.

Lelkes, Y. (2019). Policy over party: comparing the effects of candidate ideology and party on affective polarization. *Political Science Research and Methods*, 1-8.

Levendusky, M. S. (2018). Americans, not partisans: Can priming American national identity reduce affective polarization? *The Journal of Politics*, 80(1), 59-70.

Levitsky, S., & Ziblatt, D. (2018). *How democracies die*. Broadway Books.

Lipset, S.M. (1960). *Political Man*. Anchor Books.

Martherus, J. L., Martinez, A. G., Piff, P. K., & Theodoridis, A. G. (2019). Party Animals? Extreme Partisan Polarization and Dehumanization. Available online in *Political Behavior*.

Mason, L. (2015). "I disrespectfully agree": The differential effects of partisan sorting on social and issue polarization. *American Journal of Political Science*, 59(1), 128-145.

Mason, L. (2016). A cross-cutting calm: How social sorting drives affective polarization. *Public Opinion Quarterly*, 80(S1), 351-377.

Mason, L. (2018). *Uncivil agreement: How politics became our identity*. University of Chicago Press.

Mason, L., & Wronski, J. (2018). One tribe to bind them all: How our social group attachments strengthen partisanship. *Political Psychology*, 39, 257-277.

Mudde, C. (2007). *Populist radical right parties in Europe*. Cambridge University Press.

Mummolo, J., & Nall, C. (2017). Why partisans do not sort: The constraints on political segregation. *The Journal of Politics*, 79(1), 45-59.

Nuesser, A., Johnston, R., and Bodet, M.A. (2014). The Dynamics of Polarization and Depolarization: Methodological Considerations and European Evidence. Presented at *APSA* Annual Meeting.

Pierce, D. R., & Lau, R. R. (2019). Polarization and correct voting in US presidential elections. Electoral Studies, 60, 102048.

Reiljan, A. 'Fear and loathing across party lines' (also) in Europe: Affective polarisation in European party systems. Forthcoming in *European Journal of Political Research*.

Roccas, S., & Brewer, M. B. (2002). Social identity complexity. *Personality and Social Psychology Review*, 6(2), 88-106.

Rogowski, J. C., & Sutherland, J. L. (2016). How ideology fuels affective polarization. *Political Behavior*, 38(2), 485-508.

Selway, J. S. (2011). The measurement of cross-cutting cleavages and other multidimensional cleavage structures. *Political Analysis*, 19(1), 48-65.

Strickler, R. (2018). Deliberate with the enemy? Polarization, social identity, and attitudes toward disagreement. *Political Research Quarterly*, 71(1), 3-18.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In: W. G. Austin & S. Worchel. *The social psychology of intergroup relations*. Brooks/Cole, 33–47.

Tappin, B. M., & McKay, R. T. (2019). Moral polarization and out-party hostility in the US political context. *Journal of Social and Political Psychology*, 7(1), 213-245.

Van der Meer, T., van der Kolk, H., & Rekker, R. (2017). *Aanhoudend wisselvallig: Nationaal Kiezersonderzoek 2017*. Available at https://research.utwente.nl/files/84152682/Aanhoudend_Wisselvallig_NKO_2017.pdf

Wagner, M. 2017. Affective Polarization in Multiparty Systems. Presented at the *European Political Science Association General Conference*.

Ward, D. G., & Tavits, M. (2019). How partisan affect shapes citizens' perception of the political world. *Electoral Studies*, 60, 102045.

Webster, S. W., & Abramowitz, A. I. (2017). The ideological foundations of affective polarization in the US electorate. *American Politics Research*, 45(4), 621-647.