Affective polarization and social sorting: a comparative study

Eelco Harteveld
University of Amsterdam

Introduction

Politics seems increasingly tribal and divisive. This has spurred scholarly interest in the phenomenon of *affective polarization*: hostility towards compatriots with an opposing political identity (Iyengar et al. 2012; Iyengar & Westwood 2015; Mason 2015; Mason 2018; Abramowitz & Webster 2016; Iyengar et al. 2018; Wagner 2017; Reiljan 2019; Adams et al. 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 (Westwood et al. 2015; Wagner 2019; Gidron et al. 2019; 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 2019).

While present across the globe, the strength of affective polarization varies widely across time and space (Reiljan 2019). This shows that mutual dislike across ideological camps is not just a 'sign of the times', but rather something that thrives under certain circumstances but not others. Which circumstances are these? Research is only beginning to understand the micro- and macro-level mechanisms behind affective polarization, as it struggles to bridge the accumulated insights from the United States to an understanding of patterns across the world. Gidron et al. (2019a) therefore call for more *comparative* research on affective polarization. This article aims to contribute to this effort by investigating the role of social sorting: the overlap of political identities with other (non-political) fault lines in society.

In the most general sense, affective polarization involves perceiving citizens with different political views as a socially distant 'outgroup' (Tajfel 1979). This, in turn, leads people to attribute negative traits to and discriminate against this outgroup. Given that affective polarization revolves around social identities, an inquiry into its origins might well start by asking *what it is about social identities that reduce outgroup tolerance*. Social Identity Theory (Tajfel & Turner 1979) suggests that a key role is played by the way a social identity relates to *other* social identities. I follow the work of Mason (2015; 1016; 2018) and posit the general expectation that if ideologically like-minded individuals in a society *also* share multiple other non-political identities – i.e., if they are *socially sorted along political lines* – they become less tolerant towards those with divergent views. The reason is simple: 'complex' social identities create more tolerance towards outgroups than overlapping ones (Brewer and Roccas 2002). If right-wing citizens in a given society are *also* consistently (e.g.) lower educated, rural residents, ethnic majority members, and religious, while left-wing citizens have a consistently opposing profile, we can state that ideology *overlaps* with these socio-demographic identities. In that case ingroups and outgroups will look highly distinct and internally homogeneous, fostering negative outgroup affect and behavior (idem). Indeed, Mason (2015; 2016; 2018) shows that Americans with aligned religious, racial and partisan identities show more hostility towards partisan outgroup, regardless of their ideological extremity.

In this article I aim to take these insights from the United States and investigate it on a broader set of cases. I do so with two complementary studies that combine the aggregate and individual level, aiming to bridge the empirical and theoretical gap between 'macro patterns' and 'micro mechanisms'. First, building on measures of affective polarization created by Reiljan (2019) and Wagner (2017), I analyze CSES[1] data on 47 countries between 1996 and 2018 to investigate the relation between the cross-cuttingness of ideology with other social identities (on the one hand) and affective polarization (on the other). I show that if ideological divisions overlap with multiple other fault lines in society, affective polarization tends to be higher. I supplement this with a study of individual-level panel data from the Netherlands (2007-2018) that allows me to track the actual socio-demographic homogeneity of the discussion network of around 5000 individuals. I show that if such homogeneity increases, this is followed by a subsequent increase in affective polarization *of the same person* (but not the other way around). Cross-sectionally, I find that a homogeneous network is associated with self-reported social distance as well as hate towards people with different views. All of this shows that the extent to which political identities foster inter-group hostility depends on their interplay with non-political identities.

Theory

Below, I first describe the concept of affective polarization. In order for the concept to travel beyond two-party systems, I then propose a broadening in the conceptualization to include *ideological* as well as *partisan* identities. In the third section I elaborate why I expect social sorting to foster dislike of political outgroups.

Partisanship as a (divisive) social identity

Political and ideological preferences often constitute a *social identity* like any other. Many of us tend to see people who *agree with us* as *one of us* – members of a group we identify with. Like any other social identity (even the most trivial ones; Tajfel 1970), such identification has affective and behavioral implications. Among these are a disposition to like and favor the in-group ('us'). Under certain conditions, it can also result in a disliking and disfavoring towards the out-group ('them'). The resulting 'affect gap' towards the political outgroup is commonly called *affective polarization*. It is 'affective' because it involves an emotional reaction; it can be called 'polarization' because it divides individuals and groups.[2]

Can the (sometimes mundane) world of politics provide us with social identities? And if so, are these salient enough to *dislike* or even *loathe* others? Research in the United States has a longstanding tradition of conceiving partisanship as a social identity (Campbell et al. 1960) and many Americans report stable identities as either Democrat or Republican. 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

[1] Comparative Study of Electoral Systems, www.cses.org

[2] 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.

positively”. They show this is on the rise based on so-called ‘feeling thermometers’: Americans feel increasingly *colder* towards adherents of the opposite party. Americans are also increasingly unwilling to have social interaction 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 recently surged to 45% (above all among Democrats) in 2019 (Jones & Naijle 2019). These patterns are more than mere survey artefacts: Implicit Association Tests show that unconscious prejudice based on partisanship is (now) stronger in the US than negative affect based on race (Iyengar & Westwood 2015).

Of course, politics has always been heated. Democracy would not function properly without citizens caring passionately enough about a cause to participate politically on its behalf. However, there is reason to fear that sustained or growing affective polarization will negatively affect 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 various ideological colors (Ziblatt & Levitsky 2018; Strickler 2017). If political identities become strongly entrenched, and those voting for the ‘wrong party’ become loathed, the necessary tolerance (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).

It is crucial to recognize that affective polarization is not the same as ideological polarization – the extent to which citizens disagree about actual issues (or the extent to which they are apart on ideological dimensions). Of course, it makes sense to assume that if citizens disagree with each other vehemently, they will especially dislike each other. However, the relation between ideological and affective polarization is not straightforward. 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. (2019) find a weak or inconsistent relation between ideological polarization on affective polarization. Individual-level evidence on the relation on ideological and affective polarization is mixed (Rogowski & Sutherland 2016; Bougher 2017). All in all, it is clear that affective polarization is rooted at least partly – perhaps mostly – outside ideological disagreement.

Affective polarization beyond two-party systems

In the United States the nature of affective polarization is shaped by the dynamics of its two-party system, which provides a clear dichotomy which can straightforwardly absorb multiple identities at once – especially those of partisanship and ideology, which increasingly overlap (Baldassarri & Gelman 2008; Webster and Abramowitz 2017). In many other countries, the ballot offers numerous parties. How does the concept of affective polarization travel to such multiparty contexts?

Huddy et al. (2018) and Bankert et al. (2017) show that partisan identities constitute a salient social identity outside the US, too. Still, self-identification with a single party will often be smaller if multiple parties exists that are ideologically close. As a result, a partisan ingroup is not opposed to every possible partisan outgroup to the same extent. A green voter does probably not perceive a typical social democratic voter as a disliked outgroup, or at least not to the same extent as (s)he might consider a conservative voter – let alone a populist radical right voter. The same goes vice versa for populist radical right voters, who might dislike a green party voter more than a mainstream right voter.

I therefore argue that in such contexts affective polarization can be understood as *negative affect towards citizens with opposing political identities*. ‘Political identities’ should be understood to include ideological identities (Devine 2015), such as being ‘a liberal’ or ‘against immigration’, as well as partisan ones. Such positions offer citizens with labels and content for a social identity (Devine 2015). These can in turn provide not only an ingroup, but also identify an outgroup, which can become disliked (or sometimes worse). In a given context, ideological distance might generate more negative along some dimensions (such as cultural issues) than others.[3]

Indeed, Mason (2018) and Lelkes (2019) show that Americans, too, construct identities around ideological positions as well as partisan cues: “Americans are dividing themselves socially on the basis of whether they call themselves liberal or conservative, independent of their actual policy differences” (Mason 2018: 878). The fault line between ‘Remainers’ and ‘Leavers’ in the United Kingdom shows that ideological distance can be a source of mutual dislike even if it cross-cuts partisanship. Hobolt et al. (2018) show that the ‘opinion-based groups’ of Remainers and Leavers form salient social identities that create strong negative outgroup affect. Of course, this does not mean that partisan identities play no role in shaping affective polarization in multiparty systems. Even if party identities are generally weaker in multiparty systems (Huddy et al. 2015), a preference for one or more parties does constitute a powerful cue for people to label and identify political identities of themselves and other.

Social sorting as a source of affective polarization

What causes affective polarization to be more virulent in some contexts than others? Much academic (and public) attention has been given to the role of the information environment. ‘Echo chambers’, the result of ‘high-choice environments’ (Lelkes et al. 2017) both in the more traditional sense (e.g. cable news) and on social media, have been hypothesized to be particularly conducive to affective polarization. Indeed, while there is some evidence for this (Hutching et al. 2019), Iyengar et al. (2018) show that the evidence from the US is mixed at best[4], suggesting the media environment catalyzes rather than causes affective polarization. Another strand of research looks at the role of political elites (Iyengar et al. 2012), but this has proven difficult to operationalize across time and space.

Because affective polarization is rooted in social identities, a fruitful way to think about its origins is to consider what makes such identities more exclusive in general. *Social sorting along political lines*, or the overlap of political identities with non-political identities (from now on denoted simply as ‘social sorting’[5]), is likely to do just that. The reason is that it makes political outgroups much more identifiable and stereotypical. In a term coined by Roccas and Brewer, social sorting decreases Social Identity Complexity. Imagine a situation in which I associate a political outgroup (say, “progressives”) to typically be an outgroup in many other respects as well (say, “university graduates”, “atheists” and “city dwellers”). In that case social identity complexity can be said to be low. Reversely, if I feel that a political outgroup (“progressives”) is often on the ingroup side in other respects (“us lower educated” or “us rural folk”), this would create higher social identity complexity.

[3] Indeed, Gidron et al. (2019c) find some evidence that negative affect is stronger in dyads that are opposed on the cultural (rather than economic) dimension.

[4] These studies suggest that selection into highly partisan content is more often the result, rather than the cause, of polarization; the media diet of most Americans is on balance actually quite neutral; and affective polarization actually increased most among those least likely to gather their information online.

[5] The term ‘sorting’ has been used to denote a range of related phenomena, most importantly (in the US context) the overlap of ideological and partisan identities. This is of interest in the US case, where such overlap has been historically low until recent decades, but not the approach of this comparative paper. Rather, I am interested in the overlap of political identities – derived from either a partisan or ideological label – with non-political identities.

A more complex social identity can be expected to enhance tolerance for outgroups for several reasons (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. As a result, groups such as "progressives" and "conservatives" will appear less like a threatening homogeneous and distinct block. Furthermore, an overly strong dislike of a group of which the members share some features with you would create cognitive inconsistencies that need to be avoided. A complex social identity moreover reduces the extent to which any single identity can "[satisfy] an individual's need for belonging and self-definition" (p. 102). Indeed, stressing (in experimental settings) that outgroup members are ingroup members on another dimension decreases bias (Gaertner et al. 1993).

Of course, the fact that individuals have 'objectively' overlapping identities (i.e., have a collection of attributes that generally go together with a particular ideological position) does not necessarily mean they also perceive it as such. Roccas and Brewer (2002) and Mason (2018) therefore also study the cognitive representation of such overlap.[6] Still, the 'objective' configuration of social identities in a given society (i.e. the correlation of political and non-political divisions) is an important factor shaping individuals' social identity complexity – and thus their affective response towards outgroups.

In the US context, Mason (2015; 2016; 2018) studied the role of social sorting along political lines extensively. She shows that Republicans who are also White and Christian are much more affective polarized than Republicans with a cross-cutting identity – regardless of ideological extremity. The same mechanisms are observable among Democrats and their alignment with a secular and non-White identification. 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). Affective polarization was lower if he reminded respondents of their overarching *American* identity. It was also lower among citizens who were interviewed on the national holiday July 4th. If respondents were reminded that the *political outgroup* is a at the same time a *national ingroup* this immediately reduced perceived outgroup homogeneity.

My guiding hypothesis is therefore as follows:

social sorting along political lines increases affective polarization

Social sorting is likely vary both between (sub)national contexts and over time. After all, the extent to which ideological groups are socially sorted is affected by a complex set of social processes. For instance, 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. Processes of 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. I expect this variation to be a factor behind differences in the extent of affective polarization.

Design

[6] For instance, Roccas and Brewer show that under conditions of stress people's social identities become less complex.

If social sorting fosters affective polarization, this would show in two complementary ways: on the individual level (are persons with a complex social identity less affectively polarized?) and on the aggregate level (is affective polarization in a society stronger if ideology is strongly related to other fault lines?). In the remainder of this paper, I investigate both of these in two complementary studies.

In the first study, this relation is investigated on the aggregate level. Inspired by Segway (2011), I develop a macro-level *Cross-Cuttingness Sore* that captures the extent to which ideological divisions are independent of other important social cleavages in a society (that is, whether these are cross-cutting rather than sorted). To be sure, this measure uses ascribed identities that can be 'objectively' measured (education, income, urbanity), and which might not correspond to individuals' self-categorization. However, they constitute such important fault lines in many societies that I expect them to have some systematic bearing on citizens' subjective identities. I then use this *Cross-Cuttingness Score* to predict affective polarization, which I measure following the operationalization suggested by Reiljan (2019) and Wagner (2017), in repeated-cross-sectional CSES data on 47 countries between 1996 and 2018. This time span provides sufficient variation *within countries*, which is the most interesting source for causal claims. This design also ensures that my claims travel beyond just a small set of countries.

Still, any aggregate correlation between sorting and affective polarization would not "prove" that the expected mechanism is at work the individual level. To that end, I complement my design with the second study, which tests observable implications of the Social Identity Complexity mechanism (Brewer and Roccas 2002). I use Dutch panel data (LISS) that allows to track the characteristics of around 5000 individuals' social network over time up to ten years (2008-2018). I develop a measure of the extent to which this network consists of persons with similar non-political identities as that of the respondent. To the extent that it is, their Social Identity Complexity is likely to be lower. The panel data allows me to establish whether having more demographically 'cross-cutting' contacts, especially among their political discussion partners, reduces affective polarization. I discuss the design and results of each study in turn.

Study 1: affective polarization and social sorting in 47 countries

Do countries become more affectively polarized if they become more sorted?

Data and cases

The *Comparative Study of Electoral Systems* data consists of harmonized election studies in countries around the world between 1996 and 2018. This provides a unique collection of identical (or at least equivalent) survey items across all continents over 20 years, collected at a moment when political identities are most salient and impactful – elections. I only selected the 47 countries that are reported as "Free" by *Freedom House* over the entire period.

Operationalization

Measuring affective polarization

My dependent variable is Reiljan's (2019) *Affective Polarization Index* (API). Like the measure proposed by Wagner (2019) and Adams et al. (2019), this is based on sympathy scores towards various parties. Sympathy towards parties is not the same as affect towards *partisans*, but because of the correlation between the two ($r = 0.69$ in Iyengar et al. 2012; see also Druckman et al. 2019) it can still be used to infer affective polarization (especially differences over time or across contexts). Of course, views towards parties also pick up ideological distance, and as a result the constructed measure of affective polarization partly captures perceived ideological polarization (Wagner 2019). Fortunately, it is possible to control for ideological polarization, assuming the remaining part is mostly affective.

The *Affective Polarization Index* is measured as follows (Reiljan 2019: 5-6). In a nutshell, it calculates the distance in sympathy, from the point of view of the supporters of each party, to all other parties, weighted by the size of these parties. The formula below summarizes the procedure. First, for a partisan group (i.e. respondents who say they identify with a particular party) the average evaluation towards *all other parties* 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.

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

Measuring social sorting

Social sorting along ideological lines is based on the measure of cross-cuttingness developed by Selway (2011). The latter measure captures how well an individual's position on one cleavage *predicts* their position on another. If it does, the two identities go together; if it does not, they can be said to be cross-cutting. In more formal terms, the *Cross-cuttingness* measure reflects 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). The cross-cuttingness of two identities in a particular country at a particular election is calculated as 1 minus Cramer's V. The following equation (copied from Selway 2011: 52) formalizes this relation:

$$\textit{Cross-cuttingness}_{\textit{identity1, identity2}} = 1 - \sqrt{\left[\sum \frac{(O-E)^2}{E}\right] \Big/ nm,} \quad (2)$$

For each of the countries and elections in the dataset, I calculate the cross-cuttingness of political ideology (measured, based on 0-10 Left-Right self-identification, as "Far left" [0-2], "Moderately left" [3-4], "Centrist" [5], "Moderate right" [6-7] and "Far right" [8-10]) with a set of variables reflecting non-political divisions: income (5 quintiles), education (5 categories), urbanity (four categories), and ethnicity (ethnic majority vs ethnic minority[7]). This set contains 'classic' class, ethnic and spatial[8]

[7] This variable is derived from respondents' ethnic self-identification. Because the number and nature of answer options varies widely between contexts (which in itself affects the strength of the relationship with ideology), I opted to dichotomize this variable into majority ethnicity versus the rest.

cleavages while also capturing the education cleavage that increasingly structures both society and politics (Kriesi et al. 2008). On the downside, the reliance of Left-Right (rather than, say, a cultural and economic dimension) captures ideology only in a very general sense. This will likely downplay the actual level of sorting.

This measure is a priori agnostic about the shape of the relation between ideology and each of the other identities. For instance, in some countries, the lower educated are more often right wing, in other countries more often left-wing; in again other countries, they are often either far left or far right, but less often centrist. All of these patterns are captured by the cross-cuttingness measure: it reflects how well categories of education (or income, or urbanity) predict categories of ideology. The better it does, the lower the score on cross-cuttingness.

Of course, each of these measures consists of 'bivariate' cross-cuttingness (ideology and education; ideology and income; ideology and urbanity). The sorting mechanism assumes that *cumulative* identities are most powerful. I therefore calculate the overall *Cross-Cuttingness Score* of a context *c* as the product between the four dyads:

$$\textit{Cross-cuttingness Score}_c = \begin{array}{c} \textit{Cross-cuttingness}_{\textit{ideology, income}} \times \textit{Cross-cuttingness}_{\textit{ideology, education}} \\ \times\ \textit{Cross-cuttingness}_{\textit{ideology, urbanity}} \times \textit{Cross-cuttingness}_{\textit{ideology, ethnicity}} \end{array} \tag{3}$$

The ethnicity question is only included in part of the surveys, severely reducing the number of available cases (from 131 to 61). I therefore check whether the findings replicate with a measure that omits this identity.

Other variables

As control variables, I include other characteristics that can be expected to matter for affective polarization. Most importantly, I include *Ideological polarization on the elite level* and *Ideological polarization on the mass level*. Both of these are related to affective polarization theoretically: as discussed above, if parties are ideologically more distant, they are more likely to be disliked (see also Reiljan 2019). By controlling for the actual level of ideological polarization, I aim to 'isolate' as far as possible the truly *affective* 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).[9] *Ideological polarization on the mass level* is simply the standard deviation in left-right positions of respondents in a given year and country.

Another important control variable is the *Effective number of parties*. Gidron et al. (2019b) show that majoritarian systems experience more affective polarization than more consensual proportional systems, as the former system raises the stakes and produces more unequal outcomes. I therefore add a measure of the *Effective number of parties* as developed by Laakso and Taagepera (1979).[10]

Second, I include the *Salience of cultural issues* and *Salience of economic issues*. Gidron et al. (2019c) 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

[8] I opted for urbanity rather than region because the latter is operationalized in very diverging ways in CSES. Of course, the classic center-periphery cleavage does not fit perfectly on urban-rural differences: peripheral areas can be urban, too. However, in many countries the two correlate. Furthermore, urbanity per se increasingly structures lifestyle and opportunities (Guilley 2019).

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

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

share of the manifesto's of all parties in a given election that is devoted to cultural issues or economic issues, respectively.[11]

Countries are grouped by *Region* – for descriptive purposes only – as follows: Western Europe, Southern Europe Northern Europe, Eastern Europe, Latin-America, North-America, and Oceania.[12] The eurocentricity of this classification reflects the availability of cases, which allow to subdivide Europe but not Latin America (7 countries) or Asia (3 countries).

Design

Because both affective polarization and social sorting are measured at the level of country-election, I will perform the main analysis at the aggregated level. I am foremostly interested in variation *within countries*, *over time*. In other words: do countries become more (less) affective polarized if they become more (less) sorted? This provides most causal leverage as it corrects for potential (time-invariant) confounders at the country level. After all, even if – say – the UK is for some unrelated reason relatively affective polarized *and* sorted, we should still observe that in years in which social sorting is lower (in UK terms), affective polarization should be lower too (in UK terms). I therefore perform the main analysis as aggregated panel regressions with fixed effects for countries. (I then replicate this with a between-countries design.) In the descriptive bivariate graphs, I standardize variables within countries, thus restricting this analysis to within-country variation too.

A within-country design limits the analysis to the actual variation that exists within countries, which is probably much smaller than the variation between countries: affective polarization or social sorting does not change overnight. This design will provide conservative estimates of the true relation.

Results

We start with a descriptive figure. How has affective polarization developed in the seven regions? And how does this development compare to both ideological polarization and social sorting? Because elections take place at different times in different countries, average scores for a single year or group of years are strongly affected by individual countries. I therefore summarize the three variables for each region in the 1990s, 2000s and 2010s. Of course, as a general caveat it must be noted that each region is internally very diverse. Furthermore, the continental categories of "Latin America" or "Asia" actually consist of a relatively small subsection of these respective continents, and should therefore be interpreted with some caution. Appendix A shows the trends for each individual country.[13]

FIGURE 1. TRENDS IN AFFECTIVE POLARIZATION, IDEOLOGICAL POLARIZATION, AND SORTING; BY REGION.

[11] Coded as cultural issues were: … Coded as economic issues were: …

[12] See Appendix A1 for the classification.

[13] CSES wave 5 is published in a preliminary version. This means that not all elections since 2012 that will eventually become part of CSES are included in it.

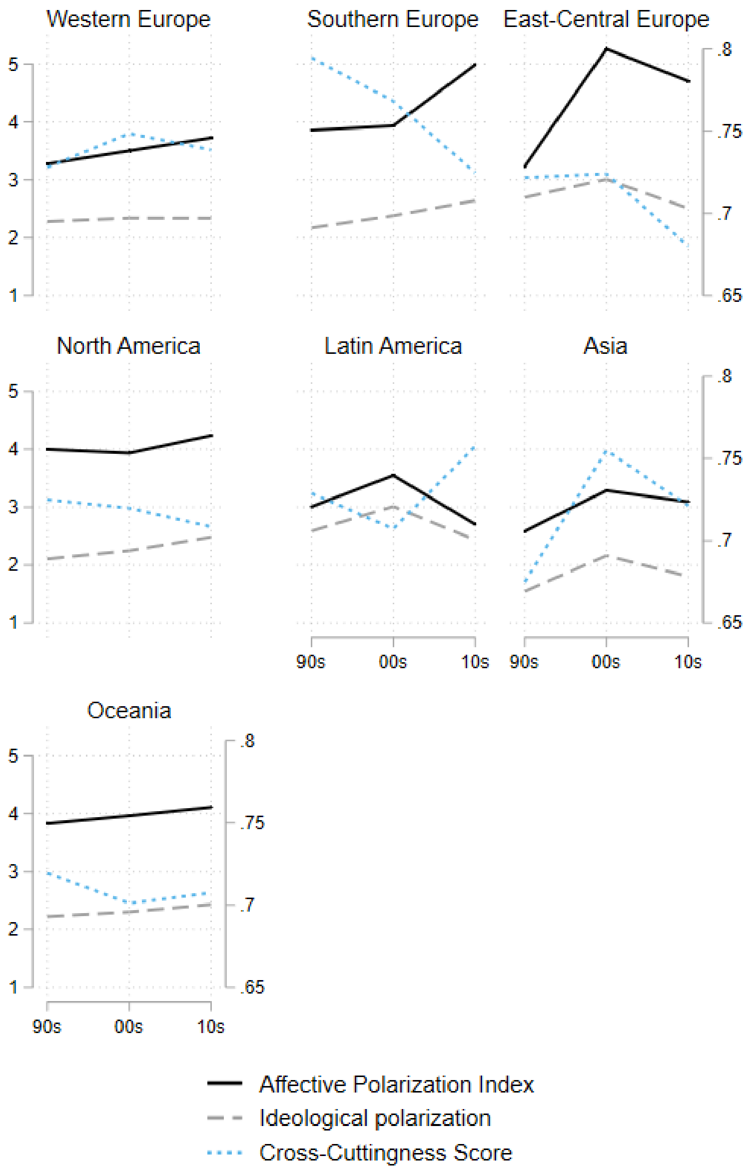


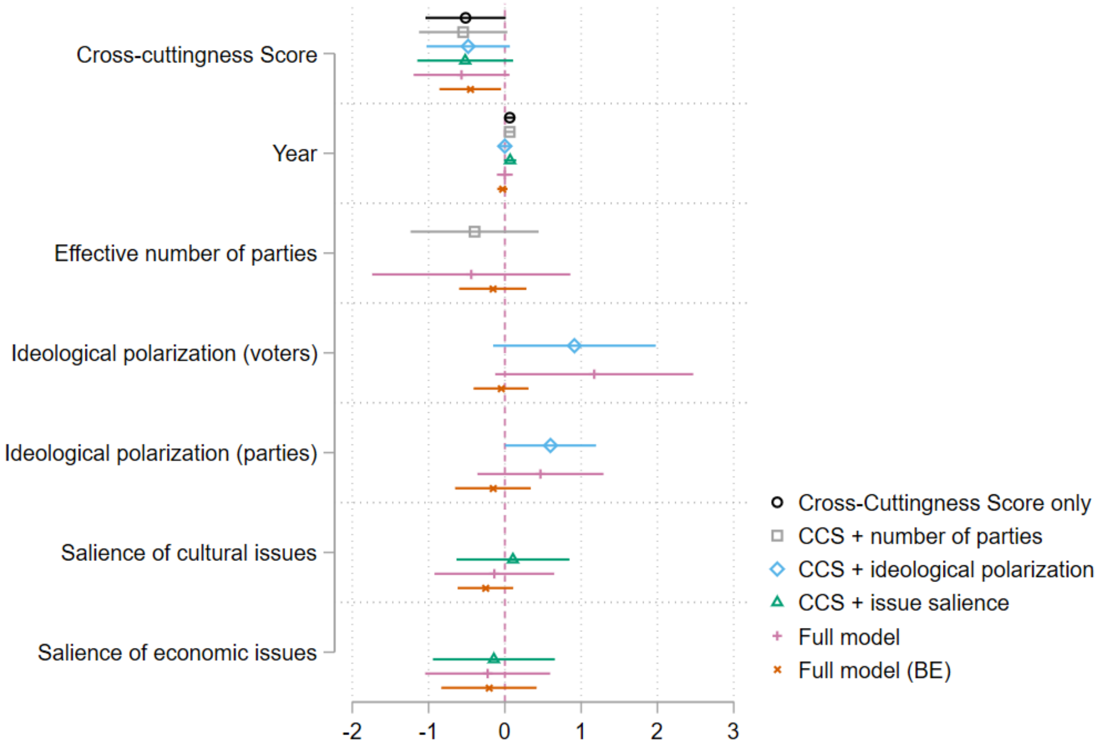
Figure 1 testifies, first of all, that affective polarization varies across the world. As reported by Reiljan (2019), Southern Europe and East-Central Europe show the highest scores (especially in later years), followed by North America (the US more so than Canada) and Oceania. Western Europe, Latin America and Asia have somewhat lower levels. Across the (very general) board, there is a weak trend of increasing affective polarization across the entire period, but not everywhere (Latin America being an exception) and not linearly so. For instance, in both Asia and East-Central Europe, affective polarization appears to have been stronger in the 2010s than in recent years. All of this confirms that affective polarization has particular trajectories in different part of the world.

On the very aggregate level, the relation between ideological polarization and affective polarization appears to be (somewhat) positive: they often rise and fall in tandem, although far from perfectly. For instance, Western Europe has a mild increase in affective polarization but stable ideological polarization. In Southern Europe, affective polarization exploded in the last decade (possibly after the economic crises), while the increase in ideological polarization was relatively more modest.

The relation of affective polarization with the *Cross-Cuttingness Score* appears mixed. In most regions, the *Cross-Cuttingness Score* decreases between the 90s and nowadays, or at least since the 10s. Latin America and Asia are clear exceptions: here, societies are less sorted than in the 90s.[14] The decrease in cross-cuttingness, and thus increase in sorting, is most spectacular in Southern and Eastern Europe, which are now the most sorted regions together with North America and Oceania. As mentioned before, we need to keep in mind that countries vary within in each of these regions. This makes it important to put the patterns to a more formal test, which we now turn to.

Figure 2 shows regression coefficients of various regression models. As mentioned, each model (expect the last) contains fixed effects for countries, and the *Affective Polarization Index* was standardized within each country.[15] To see whether different factors predict affective polarization between countries than over time, the last model is a *between-effects* (BE) model that only includes variation *between* countries rather than *within* countries.

FIGURE 2. REGRESSION COEFFICIENTS OF SIX MODELS EXPLAINING *AFFECTIVE POLARIZATION INDEX*



Note: the first five models include fixed effects; the sixth model is a between-effects model. All variables were standardized.

To test the robustness of the effect of the *Cross-Cuttingness Score* (and to observe possible mediation) each subsequent within-country model introduces a set of control variables in turn. The fifth model contains all variables. Each model contains the variable *Year* to capture possible trends in the dependent variable. All the independent variables to allow for some comparability between the different explanatory variables. Appendix X presents a replication with a *Cross-Cuttingness Score* measure without the ethnic majority component (yielding more observations).

[14] This could be a substantive pattern, but also an artifact if Left-Right self-positioning (an important ingredient in the measure) captures ideological differences less validly in these regions.
[15] This is useful because the variance in API correlates with its level.

I start by discussing the first five models – the within-country models. The first model (black circles) shows a significantly negative effect, in line with the hypotheses. In the subsequent models, under control for the other covariates, this effect remains very stable in size. This confirms the theorized pattern: the more ideology cross-cuts with other identities, the less citizens dislike ideological opponents. The effect size is substantial: a standard deviation decrease in cross-cuttingness (i.e. a standard deviation increase in sorting) is predicted to increase affective polarization by 0.48 to 0.51 standard deviations. To reiterate, this model is likely conservative because it only includes variation within countries.

In every model there is a significantly positive effect of *Year*. This means that, across the board and net of all other explanatory variables, there is a small but robust upward trend in affective polarization of about 0.04 to 0.06 standard deviations each year. This suggests that still other factors than those studied here are creating the (mild) increase in affective polarization.

Of the other variables, the strongest predictor is ideological polarization on the level of parties and (especially) voters (in line with Reiljan 2019). This makes sense: the *Affective Polarization Index* will partly capture perceived ideological differences. The fact that ideological polarization among *parties* also predicts affective polarization (if less robustly so) confirms that the behavior of elites impacts affective polarization in the public at large (Iyengar et al. 2018).

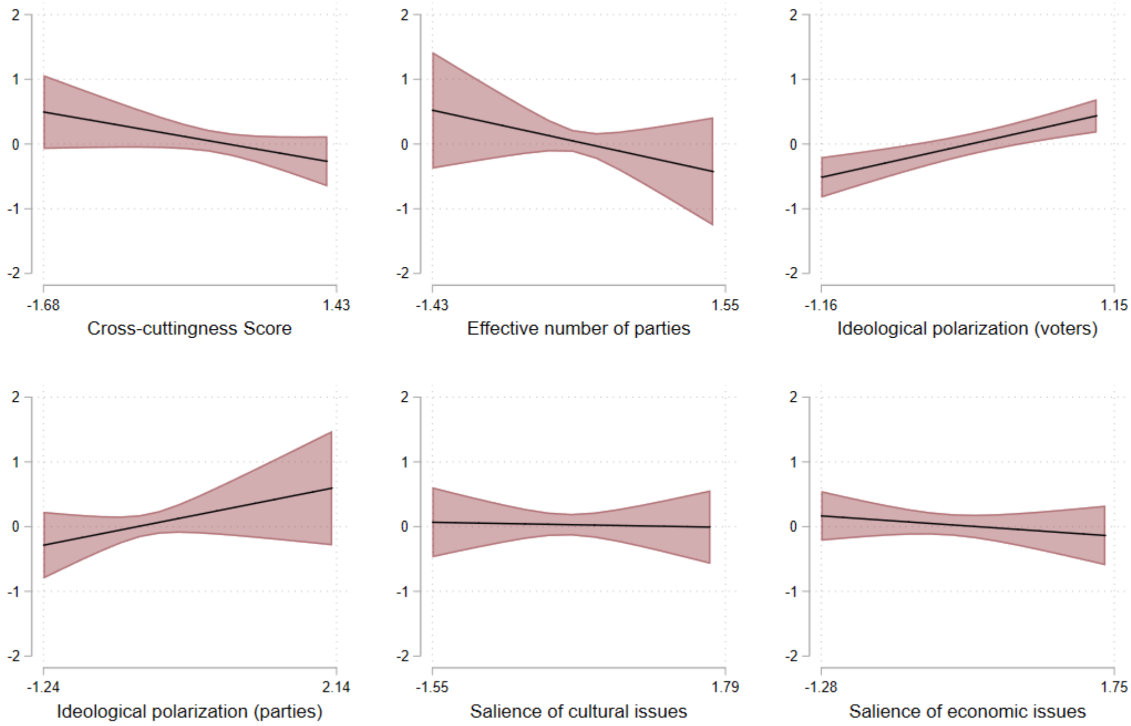
There is descriptive but insignificant relation with the *Effective number of parties*. The more parties there are in a system, the lower affective polarization appears to be. This effect is more substantial and significant in the replication on the broader set of cases (Appendix X). This pattern might reflect the more entrenched ideological and moral dichotomies that exist in two- or three-party systems, as well as the norms fostered by such systems. The salience of economic or cultural issues does not appear to affect the *Affective Polarization Index*, with or without the full set of covariates.

The very last model only relies on variation between countries.[16] The results are very similar to the within-country full model, with the exception that differences in ideological polarization between countries do not predict the level of affective polarization in those countries. This suggests that ideological polarization is perceived relate to past levels, rather than to the situation abroad. The effect of the *Cross-Cuttingness Score* is comparable with that in the within-country specification, if perhaps somewhat smaller (especially in the replication in Appendix X). This suggests that sorting, too, is better at explaining dynamics of affective polarization over time than across the world.

Returning to the within-country model, Figure 3 below visualizes the effect size by showing the predicted *Affective Polarization Index*, standardized by country, for values of each independent variable (plotted from the 5th to the 95th percentile). It confirms that the effect of social sorting is of about the same magnitude as that of ideological polarization.

FIGURE 3. PREDICTED SCORE ON THE AFFECTIVE POLARIZATION INDEX ACROSS VALUES OF THE INDEPENDENT VARIABLES

[16] In the between-country model the *Affective Polarization Index* is standardized across the whole sample, rather than within countries, because otherwise the average score in each country over time is (around) zero, leaving no variation to compare.



Robustness[17]

The analysis above relies on within-country variation only, thus decreasing the number of potential confounding factors. It could be argued that an even more convincing causal claim would consists of showing temporal succession (i.e. affective polarization increasing the election *after* social sorting increased). On the other hand, the period between elections – up to four or five years – is rather long, and it is a priori uncertain whether social sorting plays its role with such a delay. To investigate the role of temporal order, I estimated two models: (1) a model with a lagged dependent variable and lagged independent variables (i.e. explaining API by its score at the previous election as well as by the scores on the independent variables at the previous election); and (2) a first-differences model (i.e. testing whether the *change* in API is predicted by the *change* in the independent variables). A model with lagged effects shows no significant effect, and even descriptively in the other (positive) direction. Perhaps the effects play out rather quickly after social sorting increases. A first-difference model does suggest a borderline significant negative effect ($p = 0.08$).

Secondly, I replicated the analysis *on the individual level* using the measure of affective polarization developed by Wagner (2017) which is also used in Study 2 below. In a nutshell, this measure consists of a weighted standard deviation in a respondents' sympathy towards all parties (that is, weighted by each party's vote share). In a multilevel model (with a random intercept for countries), with the same predictors as the full model above plus individual-level controls for education, age, gender, and ideological position, the *Cross-Cuttingness Score* has a significant negative effect on this alternative affective polarization measure, too ($p < 0.001$).

Study 2: tracking the social network of individuals over time

[17] Results available on request.

Data, design & 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 socio-demographic homogeneity of the discussion network, as well as their affective polarization, of thousands of individuals, spanning over 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 with a more socio-demographically homogeneous discussion network (and the associated lower social identity complexity) are more affectively polarized. In this study, I do not know the ideological content that is distributed in this discussion network. However, this does not matter for the mechanism of social sorting: the point is not whether people are confirmed *ideologically* by everybody in their network (which would amount to the 'echo chamber' mechanism), but rather whether their lower social identity complexity reduces tolerance towards ideological outgroups.

The Netherlands experiences only weak affective polarization according to Reiljan (2019). This makes it a less likely case to explain its variation with sorting. To the extent that I can, it is likely that effects are at least as pronounced in more affectively polarized contexts. The source of the data is the *Longitudinal Internet Studies for the Social sciences* (LISS) panel[18], which is recruited based on a population-representative sample and runs since 2007. Its respondents answer a battery of questions on politics as well as their social network yearly. 50% of the respondents has non-missing responses to the relevant questions in at least 3 waves; 25% in at least 7 waves; and 5% in all waves (on average 4.2 waves). This provides enough variation in social networks and affective polarization. The number of nonmissing observations is around 5,000 each wave.

Operationalization

Because the unity of analysis is the individual, Reiljan's (2019) country-level *Affective Polarization Index* is not a feasible indicator. Instead, affective polarization *of an individual* is measured using Wagner's (2019) 'Weighted Affective Polarization' measure. This reflects a respondent's spread in sympathy scores towards all parties, weighted by the size of each parties. This is achieved by taking the sympathy towards a 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 therefore[19]

$$\textit{Affective Polariation}_i = \sqrt{\Sigma_{p=1}^{P} v_p(like_{ip} - \overline{like_i})^2} \quad (4)$$

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

$$\overline{like_i} = \Sigma_{p=1}^{P}(v_p * like_{ip}) \quad (5)$$

[18] www.lissdata.nl

[19] Equation copied from Wagner (2019).

I replicate the analysis with two alternative measures of an individual's affective polarization that were collected specifically for this purpose around the time of the last wave. While the one-shot measurement rules out any dynamic analysis, it does allow to see whether the correlations at least replicate cross-sectionally with alternative operationalizations. The first alternative measure is *Social distance* towards political and ideological outgroups (measured by the 'marriage question', ranging from 1 ["would be very happy] to 7 ["would be very unhappy"] if a child would marry the outgroup member[20]). Second, the extent to which the respondent agrees to the statement that (s)he "started to hate people because of their views" (*Outgroup hate*, ranging from 1 ["Strongly disagree"] to 7 ["Strongly agree"]).

I base my measure of *Network Homogeneity* on respondents' self-reported social network. LISS asks respondents to mention up to 5 closest contacts (people they "discuss important things with"). Subsequently they are asked to describe these persons in various terms. The homogeneity of the network depends on the extent to which these 5 people differ from the respondent in terms of gender (male, female), age (15-24; 25-34;34-44;45-54; 55-64; 65+), education (primary; lower secondary; higher secondary; lower vocational; higher vocational; university), and immigration background (1st/2nd generation versus not). To this end, I recoded all these variables to range from 0 to 1. For each variable I then take the *average absolute distance* between the respondent and each network member. For instance, if the respondent is male (score 1 on gender), and *all* the discussion partners are female (score 0 on gender), the average distance in terms of gender is 1. If two discussion partners are male, the average distance is 0.60, etcetera. For education, which is measured in six categories, the distance towards each network member can be 0, 1/6th, 2/6th, etc.[21] The composite measure *Network Homogeneity* for all variables is simply the *product* of the homogeneity scores of gender, age, education and immigration status (thus assuming that cumulative homogeneity has more impact than the sum of it). For individual *i*, and five network members m_1 to m_5, the equation therefore reads:

$$\textit{Network Homogeneity}_i = \begin{array}{l} \sum_{m=1}^{5}|gender_i - gender_m| \, * \, \sum_{m=1}^{5}|age_i - age_m| \, * \\ \sum_{m=1}^{5}|education_i - education_m| \, * \\ \sum_{m=1}^{5}|imm.status_i - imm.status_m| \end{array} \qquad (6)$$

Unfortunately, immigration status is only available since 2011. I therefore performed the analysis for both the entire period (without immigration) and the waves since 2011 (with immigration).

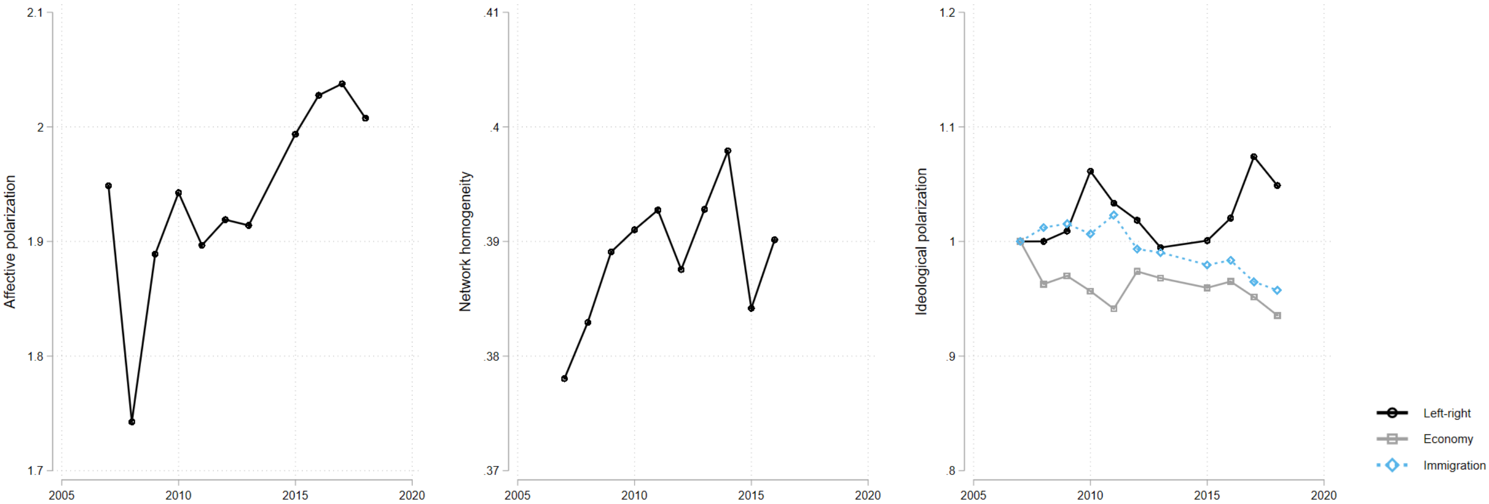
I expectthat homogeneity in a network will increase affective polarization especially if people discuss political matters with these people group. After all, this sensitizes them (implicitly or explicitly) to the fact that ideological and other identities overlap. I therefore create a measure of *Political Talk in Network* that consists of the mean score of the following question over all network members: "How often do you talk about politics with this person?" (options ranging from "Not at all" [0] to "Every day" [7]). As I will discuss below, I also create an alternative *Network Homogeneity* measure in which I weight the each respondent by the extent to which politics is discussed with that person.

Results

[20] The three outgroups were: somebody with a different view on refugees; somebody with a different view on welfare; and somebody that votes for the least-liked party of the respondent.
[21] This measure thus assumes equal distances between each education category, which is of course unlikely. This will introduce noise to the measurement.

Descriptives
Figure 4 shows aggregated trends in *Affective polarization*, *Network Homogeneity*, and ideological polarization. The latter is measured by the standard deviation, across all respondents in a wave, in responses on (1) a 0-to-10 Left-Right self-placement scale, (2) an item on immigration ("Immigrants can retain their own culture" [1] to "Immigrants should adapt entirely to Dutch culture" [5]", and (3) an item on redistribution ("Differences in income should increase" [1] to "Differences in income should decrease" [5]). Because the answer scales differ, the nominal standard deviations are very difficult to compare in absolute terms. I standardized each item and centered it around the score in the first wave. This way at least the *trend* is comparable.

FIGURE 4. TRENDS IN AFFECTIVE POLARIZATION, NETWORK HOMOGENEITY, AND IDEOLOGICAL POLARIZATION; BY WAVE.

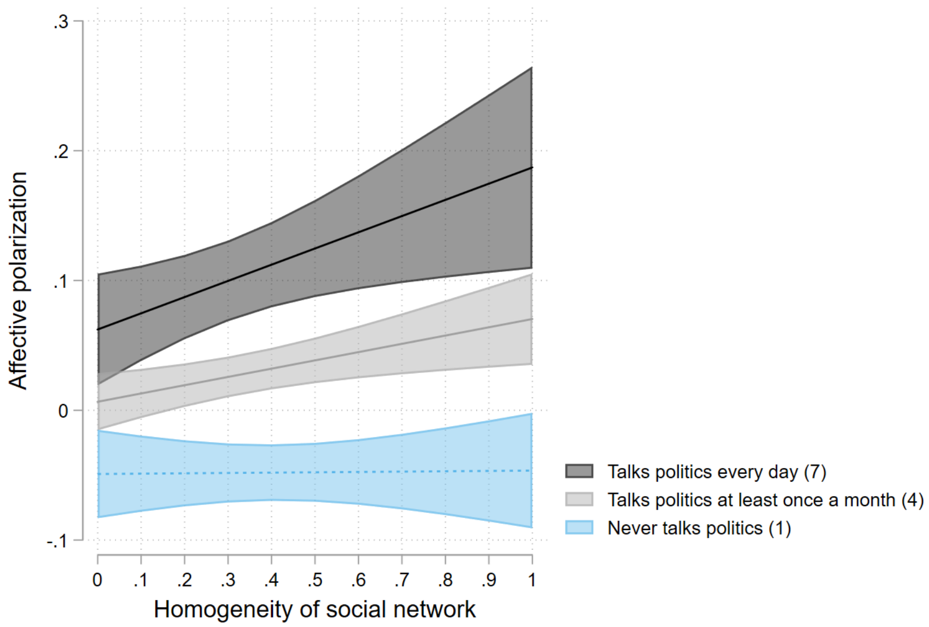


The right-most panel of the figure does not provide evidence for a strong increase in ideological polarization across the period. The standard deviation in left-right positioning has increased after 2015, but had already experienced a similar peak around 2010. The two peaks suggest that periods of heated contestation over the Great Recession and the refugee crisis create divergence among the audience. However, the trends for the item about immigration and the economy suggest no such issue-specific peaks. If anything, polarization over redistribution *lowered* after 2008, as did polarization over immigration after 2016. The trend in *Affective Polarization* suggests an upward trend after 2013. The very low score in 2008 is puzzling, and might be an artefact of changes in panel composition. Discounting that data point, affective polarization has been roughly stable until 2013 and increased somewhat since. *Network Homogeneity* seems to paint a very similar picture. Again it looks relatively stable, with (again) perhaps a small upward trend since the early century. I now turn to the individual-level correlations, which provides the most relevant test.

Multivariate analysis
I start with a regression model in which I predict a respondent's *Affective Polarization* by their *Network Homogeneity* (calculated without immigration status, thus spanning the entire panel). The model controls for gender, age, education and left-right position. As mentioned above, I expect the effect of homogeneity to differ depending on the extent to which politics is discussed in this network, so I include an interaction with *Political Talk in Network*. This interaction is in the expected

direction although significant at the 10% level. Figure 5 visualizes the results of this regression, of which the full table is available in Appendix B.

FIGURE 5. PREDICTED AFFECTIVE POLARIZATION BASED ON HOMOGENEITY OF SOCIAL NETWORK AND POLITICAL TALK IN NETWORK



The analysis shows, first of all, that people who discuss politics more often are generally more affectively polarized. For them, their political social identity is more salient. Furthermore, it becomes immediately clear that a more homogeneous social network is associated with more affective polarization (up to a quarter of a standard deviation more), though only among those who discuss politics with their network to at least some degree.

Still, the correlation could be brought about by reverse causality: affectively polarized people selecting themselves in more homogeneous networks. To find out, I use panel regression to gauge the temporal order of cause and effect. Because the LISS waves are conducted yearly, the gap between waves is not so large as in CSES, making it more plausible to observe a delayed effect.

Because the figure above suggested that network homogeneity matters more for those who talk about politics, I proceed using a *Weighted Network Homogeneity* indicator. This variable weights each network partner by the extent to which the respondent talks politics with her or him, rescaled to 0 (does not talk about politics) to 1 (talks about politics every day). This way, frequent political discussion partners feature more prominently in the *Weighted Network Homogeneity* score. The benefit of a single indicator is that it better allows for panel regressions.

TABLE 1. REGRESSION MODEL WITH LAGGED DEPENDENT VARIABLES.

	Affective Polarization *b/se*	**Weighted Network Homogeneity** *b/se*
Affective Polarization (lagged)	0.410***	0.008
	(0.006)	(0.006)
Weighted Network Homogeneity (lagged)	0.045**	0.372***
	(0.015)	(0.016)
Ideology (lagged) (ref: Far left)		
- *Moderate left*	-0.195***	-0.043*
	(0.018)	(0.018)
- *Center*	-0.350***	-0.124***

	(0.020)	(0.020)
- Moderate right	-0.317***	-0.079***
	(0.018)	(0.019)
- Far right	-0.219***	-0.082***
	(0.021)	(0.022)
Political Talk in Network (lagged)	0.017+	-0.037***
	(0.010)	(0.011)
Age (ref: 15-24)		
- 25-34	-0.073**	-0.085**
	(0.027)	(0.027)
- 34-44	-0.071**	-0.035
	(0.026)	(0.027)
- 45-54	-0.036	0.005
	(0.025)	(0.026)
- 55-64	0.007	0.048+
	(0.024)	(0.025)
- 65+	0.053*	0.01
	(0.024)	(0.025)
Female	0.005	-0.143***
	(0.012)	(0.013)
Education (ref: primary)		
- L-Secondary	0.009	0.098***
	(0.025)	(0.027)
- H-Secondary	0.090**	0.194***
	(0.028)	(0.029)
- L-Vocational	0.038	0.147***
	(0.026)	(0.028)
- H-Vocational	0.131***	0.249***
	(0.025)	(0.027)
- University	0.224***	0.321***
	(0.030)	(0.032)
Year	0.031***	0.004+
	(0.002)	(0.002)
Intercept	-61.898***	-6.912+
	(3.298)	(3.916)

*+ p<0.10 * p<0.05 ** p<0.01 *** p<0.001*

In Table 1, *Affective Polarization* and the *Weighted Network Homogeneity* are predicted score in turn. Control variables include ideology, average political discussion in the network (higher scores indicate less talk about politics), age, gender, education, and year.[22] The model includes random effects for respondents.[23] In the second column of Table 1, I predict an individual's score on *Affective Polarization* by, first of all, the lag in (i.e. the score in the previous wave of) the same variable. I furthermore predict it by her or his lagged *Weighted Network Homogeneity* score. It turns out that this significantly ($p < 0.01$) predicts affective polarization. This is also the case in a replication that included immigration status and thus runs from 2011 onwards (see Appendix B2). This even somewhat strengthens the effect (from b=0.41 to b=0.44).

In the third column, *Weighted Network Homogeneity* is the dependent variable. This turns out *not* to be predicted by the lagged score on *Affective Polarization*. This finding makes it more likely that the pattern emerging in Figure 5 did not reflect self-selection by affective polarized people, but instead a real effect of network homogeneity.

FIGURE 6. PREDICTED AFFECTIVE POLARIZATION ACROSS VALUES OF VARIOUS VARIABLES

[22] I only lag the substantive variables of interest that are actually time-variant. I repeated the analysis with dummies for each year; this yielded the same results.
[23] A fixed effects specification – which controls for all unobserved time-invariant factors – provides an effect size of almost equal size (lagged effect of network homogeneity of b=0.047).

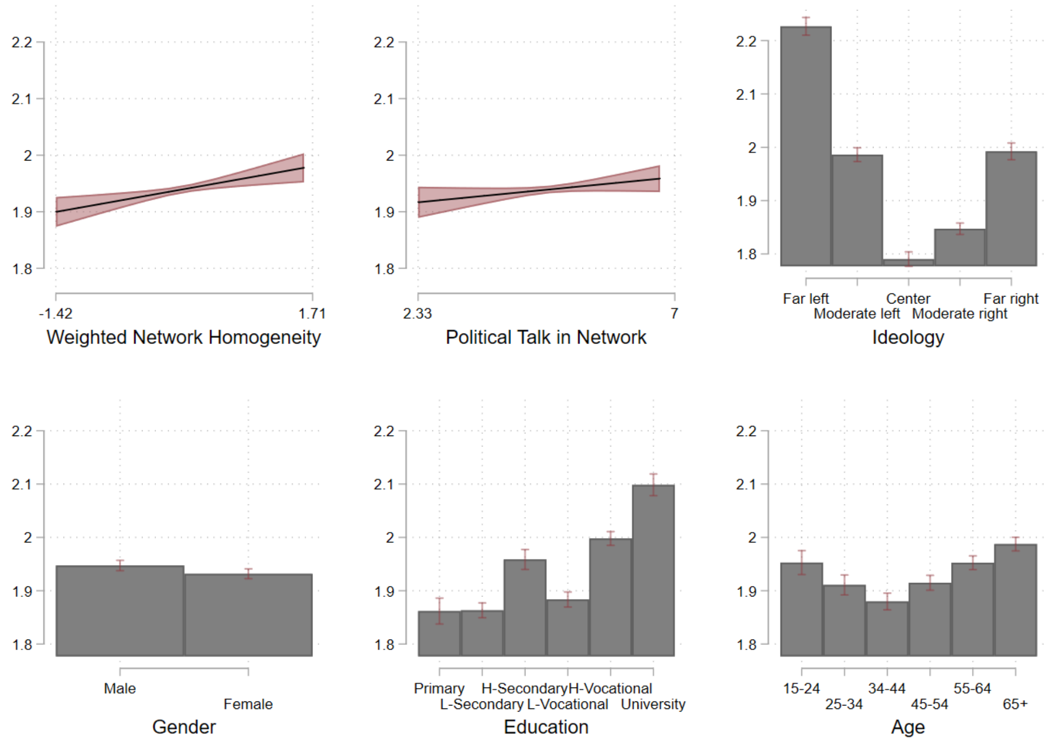


Figure 6 above summarizes the effect size of (lagged) network homogeneity, as well as the other variables, on affective polarization.[24] Ideology is the most important predictor. It should be kept in mind, however, that the variation in ideology mostly occurs *between* people, as ideology is very stable. The effects of ideology in a within-respondent model is much smaller: about the magnitude of *Network Homogeneity* (of which the effect size remains similar). Regardless of specification, those at the ideological extremes are most affectively polarized, compared to centrists or moderates. Citizens with radical views are probably especially committed to their ideological identity, and they experience a high ideological distance towards parties on the other side of the spectrum. In general, the left is more affectively polarized than the right. They might react most strongly to the presence of a populist radical right party, which creates most negative affect (Gidron et al. 2019c).

Education (which also varies virtually only between respondents) is a good predictor, too. Across the board, the higher educated are more affective polarized than the lower educated. Political identities are likely more often salient among the higher educated. Age shows a U-curved pattern: the young and the old are most affectively polarized, compared to those in the middle. However, the 10-year panel does not allow to distinguish age and cohort effects.

Alternative dependent variables

I finish this section by looking at two alternative dependent variables: *Social distance* and *Outgroup hate*. The regression analyses in Appendix B3 show that *Network Homogeneity* significantly predicts both variables (p<0.001). Respondents with more homogeneous networks are less likely to accept social interaction with an ideological outgroup member and more likely to state that they started to hate others because of their views. This furthers strengthens confidence in the conclusions of the panel regression above.

[24] Continuous variables are plotted between the 5th and 95th percentile; categorical variables across all observed values.

Conclusions

It is an old insight that cross-cutting cleavages can decrease social tensions (Lipset 1960). Building on the research of Mason (2015; 2016; 2018), I showed that societies become more affective polarized if ideology overlaps with other, non-political divisions. This effect is substantive: of the same magnitude as that of ideological polarization. I furthermore show, using Dutch panel data, that citizens whose network becomes more socio-demographically homogeneous become more affectively polarized (but not the other way around). It is telling that this mechanism is visible even in a country with very weak affective polarization (Reiljan 2019). Both studies confirm that the hostility fostered by political identities depends on its interplay with non-political identities.

(…)

Does this mean that social sorting is the key factor behind the increase in affective polarization? First of all, there is no such thing as *the* increase in either social sorting or affective polarization: I found that the trends vary substantially between countries. However, in many contexts both are on their way up. Furthermore, even if sorting is not increasing among a large share (even a majority) of citizens, if an important subset of society – probably the politically most engaged – *does* find itself increasingly sorted, then the overall result can still be an increase in affective polarization in those parts of society that shape the public and political debate. An important next step for this field would be to explore who is most 'at risk' of becoming affectively polarized.

Another important avenue for further research is to get closer to the ideological content of a person's network. This would allow to better distinguish the mechanism of 'echo chambers' (being confirmed in your views by everybody around you) from the sorting mechanisms. Such data would also allow to explore whether the cultural dimension is more divisive than the economic (due to its heightened moral resonance; Graham et al. 2008). I did not find that the salience of cultural issues matters, but Gidron et al. (2019c) did find that cultural issues are more divisive.

If social sorting indeed bolsters affective polarization, as this study suggests, high levels of affective polarization might eventually reinforce social sorting in turn. I did not find evidence for this on the short term in Study 2, but nonlinearities can be expected here: if dislike of ideological opponents reaches a certain cutoff point, people might start to select themselves into more homogeneous networks. This would result in increasing overlap on the macro level, and could create a spiral in which ideological divisions get increasingly entrenched in the fabric of society. This potential endogeneity rings also true for most explanations of affective polarization, including the content of social media (or other parts of the information environment) or politicians' moralizing and Manichean discourse. These have their roots in technological changes, the rise of populist actors, and shifting political opportunity structures. Disentangling 'prime movers' from reinforcing mechanisms would help to trace down the roots of affective polarization.

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.

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.

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*.

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

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). How Ideology, Economics and Institutions Shape Affective Polarization in Democratic Polities. *Presented at APSA.*

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

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, 96, 1029-1046

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.

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.

Larcinese, V., Puglisi, R., & Snyder Jr, J. M. (2011). Partisan bias in economic news: Evidence on the agenda-setting behavior of US newspapers. *Journal of public Economics*, 95(9-10), 1178-1189.

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. *Political Behavior*, 1-24.

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.

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

Jones, R.P and M. Najle. 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).

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. (1970). Experiments in intergroup discrimination. *Scientific American*, 223(5), 96-103.

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.

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

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

Appendix A0 Regional categorization

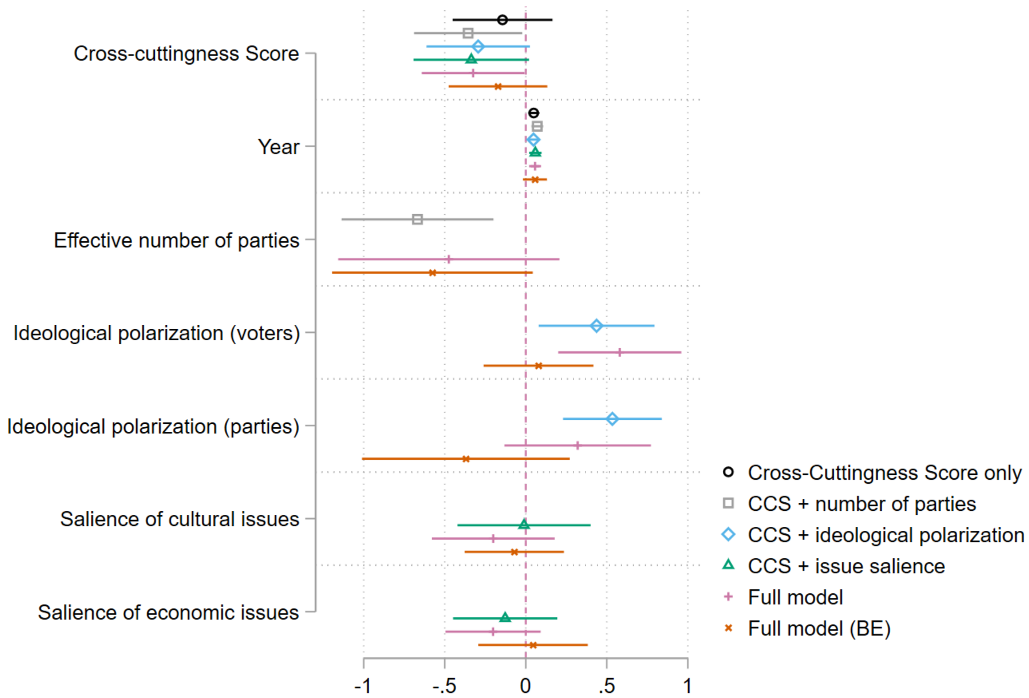
Western Europe: Austria, Belgium, Denmark, Finland, France, Germany, Great Britain, Iceland, Ireland, the Netherlands, Norway, Sweden and Switzerland. Southern Europe is Greece, Italy, Portual and Spain.
East-Central Europe: Albania, Bulgaria, Croatia, Czechia, Estonia, Hungary, Latvia, Lithuania, Montenegro, Poland, Romania, Slovakia and Slovenia.
North America: Canada and the United States.
Latin America is Argentina, Brazil, Chile, Mexico, Peru and Uruguay.
Asia is Japan, South Korea and Taiwan.
Oceania is Australia and New Zealand.
Note: Israel and South Africa unfortunately do not fit in the above categorization, nor do they suffice for a separate region. They are still included in the overall models.

Appendix A1 Trends in key variables per country

Appendix A2 Regression model behind Figure X

	fe1	fe2	fe3	fe4	fe5	be
	b/se	b/se	b/se	b/se	b/se	b/se
Cross-cuttingness Score	-0.514+	-0.546+	-0.481+	-0.520+	-0.568+	-0.453*
	(0.254)	(0.276)	(0.259)	(0.299)	(0.293)	(0.175)
Year	0.064+	0.063+	0	0.069+	-0.001	-0.032
	(0.032)	(0.033)	(0.044)	(0.038)	(0.048)	(0.029)
Effective number of parties		-0.397			-0.441	-0.158
		(0.401)			(0.606)	(0.191)
Ideological polarization (voters)			0.912+		1.171+	-0.051
			(0.505)		(0.605)	(0.156)
Ideological polarization (parties)			0.598*		0.467	-0.156
			(0.283)		(0.385)	(0.215)
Salience of cultural issues				0.106	-0.138	-0.255
				(0.352)	(0.366)	(0.158)
Salience of economic issues				-0.144	-0.225	-0.208
				(0.380)	(0.382)	(0.271)
Intercept	-127.991+	-126.939+	0.04	-137.759+	1.471	64.579
	(63.233)	(66.965)	(88.197)	(76.570)	(96.882)	(58.702)

Appendix A3 Replication of analysis behind Figure X excluding 'ethnic majority' in the CCS measure



Appendix B1 Regression behind Figure 2

	b/se
Network Homogeneity	0.145*
	(0.063)
Political Talk in Network	-0.019***
	(0.005)
Network Homogeneity X Political Talk in Network	-0.020+
	(0.012)
Age	0.031***
	(0.004)
Gender	0.033*
	(0.015)

Education (ref: Primary)	
L-Secondary	-0.070**
	(0.026)
H-Secondary	0.095***
	(0.027)
L-Vocational	-0.029
	(0.026)
H-Vocational	0.138***
	(0.026)
University	0.298***
	(0.030)
Ideology (ref: Far left)	
Moderate left	-0.368***
	(0.017)
Center	-0.598***
	(0.019)
Moderate right	-0.509***
	(0.018)
Far right	-0.221***
	(0.021)
Intercept	0.206***
	(0.048)

*+ p<0.10 * p<0.05 ** p<0.01 *** p<0.001*

Appendix B2 Replication of Table X with a network homogeneity measure including immigrant status (2011-2018)

	Affective Polarization *b/se*	**Weighted Network Homogeneity** *b/se*
Affective Polarization (lagged)	0.448***	0.003
	(0.007)	(0.007)
Weighted Network Homogeneity (lagged)	0.053*	0.536***
	(0.022)	(0.024)
Ideology (lagged) (ref: Far left)		
- Moderate left	-0.186***	-0.02
	(0.023)	(0.024)
- Center	-0.325***	-0.115***
	(0.026)	(0.027)
- Moderate right	-0.287***	-0.066**
	(0.024)	(0.025)
- Far right	-0.251***	-0.074**
	(0.027)	(0.028)
Political Talk in Network (lagged)	0.026+	0.053**
	(0.015)	(0.017)
Age (ref: 15-24)		
- 25-34	-0.109**	-0.063
	(0.038)	(0.039)
- 34-44	-0.093*	-0.059
	(0.037)	(0.038)
- 45-54	-0.076*	0.013
	(0.036)	(0.036)
- 55-64	-0.053	0.062+
	(0.035)	(0.035)
- 65+	0.01	0.023
	(0.035)	(0.035)
Female	0.005	-0.112***
	(0.016)	(0.017)
Education (ref: primary)		
- L-Secondary	0.063+	0.052
	(0.034)	(0.037)

- H-Secondary	0.113**	0.101*
	(0.038)	(0.040)
- L-Vocational	0.083*	0.089*
	(0.035)	(0.037)
- H-Vocational	0.173***	0.161***
	(0.034)	(0.037)
- University	0.273***	0.199***
	(0.039)	(0.042)
Year	0.031***	0.002
	(0.003)	(0.004)
Intercept	-61.505***	-4.443
	(5.350)	(7.210)

Appendix B3 Predicting alternative measures of affective polarization

- *Social distance* is measured by the following question: "Imagine a son or daughter of yours would get married. How would you feel about the following?". The three subitems started with "If (s)he would marry somebody who…" and were followed by a description of a potential spouse defined by (1) a position on immigration opposed to that of the respondent; (2) a position on general benefits opposed to that of the respondent; (3) voting for the party that the respondent indicated to dislike most in a previous question. The answer options ranged from 1 ("I would find that very problematic") to 4 ("I don't mind") to 7 ("I would be very happy about that").
- *Outgroup hate* is measured by the extent to which respondents agree with the statement "There are people I started to hate because of their views". Answer options ranged from 1 ("Fully disagree") to 5 ("Fully agree").

	Social distance b/se	Outgroup hate b/se
Weighted Network Homogeneity	0.106***	0.077***
	(0.016)	(0.012)
Age	0.040***	0.121***
	(0.010)	(0.008)
Gender	0.033	0.082***
	(0.030)	(0.024)
Education (ref: primary)		
L-Secondary	-0.231***	-0.376***
	(0.065)	(0.052)
H-Secondary	-0.273***	-0.033
	(0.074)	(0.059)
L-Vocational	-0.262***	-0.302***
	(0.065)	(0.051)
H-Vocational	-0.371***	-0.152**
	(0.064)	(0.051)
University	-0.327***	-0.067
	(0.074)	(0.059)
Ideology (ref: far left)		
Moderate left	-0.255***	-0.364***
	(0.050)	(0.040)
Center	-0.246***	-0.511***
	(0.052)	(0.041)
Moderate right	-0.319***	-0.509***
	(0.047)	(0.037)
Far right	-0.363***	-0.422***
	(0.054)	(0.043)
Intercept	2.628***	4.533***
	(0.103)	(0.082)