ORCID:
LR: 0000-0002-1237-7310
ET: 0000-0002-8332-7820
EH: 0000-0003-4233-6592
AH: 0009-0006-1886-0406

This is not US: measuring polarization in multiparty systems. A quasi-replication study

Luana Russo[1], Emma Turkenburg[2,3], Eelco Harteveld[4,*], Anna Heckhausen[1,5]

[1] *Maastricht University*
[2] *Wageningen University & Research*
[3] *KU Leuven*
[4] *University of Amsterdam*
[5] *Central European University*
*Corresponding author. E-mail: E.Harteveld@uva.nl

Abstract. In the last decade, affective polarization (AP) has become an increasingly salient topic in both public discourse and political science. Several different measurement instruments have been developed to empirically capture this phenomenon. With the rising interest that affective polarization is now also enjoying in Europe, it has become of the utmost importance to assess what these different measures capture, and to what extent their application travels to different contexts. In this study we test several AP measures on a student population with various European nationalities. We assess their overlap and effectiveness in mapping AP, to help future research working towards greater empirical clarity, and making informed choices on which kind of measures to include in questionnaires and data collections. The results indicate that, while different items usually produce different point estimates and sometimes different answer patterns, the measurement of affective polarization appears relatively indifferent to the choice of items.

Keywords: affective polarization, measurement, methodology, political behaviour.

INTRODUCTION

Affective polarization refers to "view[ing] opposing partisans negatively and copartisans positively" (Iyengar & Westwood, 2015, p. 691), or "hostility between rival political partisans" (Huddy & Yair, 2020, p. 1). The topic has attracted a lot of scholarly attention in the last ten years. Most of these studies are based in the United States, where affective polarization (hereafter AP) was first observed and studied. This great interest is likely due to the very detrimental potential consequences of AP for societal cohesion and democratic health (McCoy and Somer, 2019; Iyengar *et al.*, 2018; Mason, 2018b).

The idea of democracy is that different worldviews compete for citizens' consent, and peacefully alternate in response to that consent. But as Lipset (1959) noticed: "Inherent in all democratic systems is the constant threat that the group conflicts which are democracy's lifeblood may solidify to the

point where they threaten to disintegrate society" (p.83). Some degree of elite polarization may be beneficial to offer voters clear cues activating the heuristics that lead to the decision to vote (also known as *sorting*) (Russo *et al.*, 2021). Furthermore, high levels of AP in the public can increase political participation (Iyengar & Krupenkin, 2018; LeBas, 2018; Levendusky, 2010). However, as Mason (2018b) argues, the *reasons* to participate in politics matter, and high levels of mass AP might lead to increased intergroup animosity, hampering democratic processes by discouraging compromises, and even leading to an escalation of conflict. In sum, affective polarization can harm the basic principles of a well-functioning democracy (Iyengar *et al.*, 2019; Mason, 2018b; McCoy, Rahman, & Somer, 2018), reaching the solidification of democracy's blood that Lipset (1959) feared.

The detrimental consequences of AP have already urged many scholars to try to capture this phenomenon empirically, and several measurement instruments have been developed in the last decade (Druckman & Levendusky, 2019). AP is a broad concept, and so far, it has been operationalized in several ways. The attention to measurement instruments is thus critical for the development of the field. Druckman and Levendusky (2019) investigated how different measures relate to one another in the US. Their results show that the feeling thermometer, traits evaluation and trust measures are highly correlated, and that only social distance measures (that is, the willingness of interacting with other party's supporters on several levels) can be considered really different (for a more detailed review of the available operationalizations, see the section *Measurement of affective polarization* below). They also found that voters rate party elites more negatively than party supporters. In short, all three most used measures effectively capture affective polarization among Americans, and researchers can pick the most appropriate one in accordance with their research question(s).

As part of the rising scholarly attention to AP in Europe, some of these measures have also been employed in research in European contexts (e.g., Knudsen 2020 Harteveld, 2021; Harteveld, Mendoza, & Rooduijn, 2021; Kekkonen & Ylä-Anttila, 2021, Van Erkel & Turkenburg 2022). However, this raises the question of whether these measurements, which were developed and aimed to measure affective polarization in the US, work equally well in the culturally and institutionally diverse context of Europe. Our study aims to contribute to answering this question.

On the one hand, items that are applied to a different context still pick up on fundamentally similar mechanisms. A wealth of studies has convincingly shown that AP is ingrained in the fundamental human need to distinguish between in- and out-groups (Iyengar, Sood, & Lelkes, 2012; Mason, 2018b; Tajfel *et al.* 1971), and it could well be that these mechanisms are nearly universal. If so, items should travel easily to different contexts. However, prominent differences exist between the US and European settings, as well as between different European societies. Notions of (and even the very words) 'liking', 'trusting', feeling 'warm', or wanting to 'avoid' somebody are highly specific to cultural contexts. Considering that the social psychology literature recommends caution in making assumption about the functioning of attitudes across contexts (see Hogg & Smith, 2007), it is pivotal to empirically test whether this assumption holds. Recently, Gidron *et al.* (2022), following this very same argument, provided a validation for the party feeling thermometer in a multi-party system (Israel), and found that thermometer scores reflect sentiment towards party supporters, and demonstrated that they go hand-in-hand with preferences for social distance and discrimination in economic games. In this paper we follow the logic of Druckman and Levendusky (2019) and investigate how different operationalizations of AP (not only the feeling thermometer) operate vis-à-vis one another on a sample of European university students drawn from different nationalities. If items perform similarly across this sample, and similarly to the US context, it is reasonable to assume they have strong cross-cultural applicability.

The paper proceeds as follows. In the next section we offer an overview of the current AP measurement and operationalizations. Then, we briefly discuss why these measurement instruments might lead to different patterns in Europe. Finally, following Druckman and Levendusky (2019) we perform a test of how these measurements perform in respect to one another by testing them on a student sample. Our aim is to help future research making informed choices when deciding which measures of affective polarization to include in questionnaires.

MEASUREMENTS OF AFFECTIVE POLARIZATION

Measurements of AP range from measurements based on respondents' general attitudes or affect towards others to measurements assessing social distance or actual behaviour. Respondents are usually asked to rate their feelings or give their opinion on their political ingroup and outgroup, and afterwards the presence and size of a 'gap' in affect is scrutinized (Reiljan, 2020). In US-based research this generally concerns asking people

how they feel about Republicans and Democrats (Iyengar *et al.*, 2012; Lelkes, 2016). In a European context, these questions are asked about all parties (Wagner, 2021) or a selection of (the largest) parties (see e.g., Westwood *et al.*, 2018).

Before moving to more concrete operationalizations, we point out that measurements have been employed to evaluate **objects** at different levels. Most commonly, such scales have measured affect towards abstract parties (such as 'The Democratic Party' or 'Alternative for Germany'; Iyengar & Krupenkin, 2018; Mason, 2015), but they have also been used to measure affect towards leading politicians or candidates (Garrett *et al.*, 2014; Ondercin & Lizotte, 2020) whilst the comparative use of both thermometers at the same time is quite rare (Webster & Abramowitz, 2017). Although the use of items targeting elites or abstract actors is widespread, these measures have also been subject to criticism. For instance, Klar, Krupnikov, and Ryan (2018) have shown that thermometers may (at least partially) lead researchers to misinterpret disdain for a specific party with what in fact is disdain for parties per se (anti-system voters). Relatedly, Kingzette (2021) conducted an experiment which shows how citizens tend, on average, to dislike the leaders of a party more strongly than its supporters. Druckman and Levendusky (2019) also find that it is important to exercise caution in phrasing the object of affective polarization. Their research shows that asking about 'Republicans' and 'Democrats' in the abstract makes people think about elites, rather than their fellow voters. It is therefore important to specify the object of polarization that is asked about, especially since the level of affective polarization can differ strongly dependent on whether questions refer to a party, a party elite, or voters of a party (Druckman & Levendusky, 2019; Duffy, Hewlet, McCrae, & Hall, 2019; Iyengar *et al.*, 2012). Indeed, in multiparty systems too, affective polarization is increasingly measured using items asking for evaluations of supporters of parties (Harteveld 2021; Kekkonan & Ylä-Anttila 2021; Van Erkel & Turkenburg 2022).

The most commonly used affect measures are the **'like-dislike'** scale and the '**feeling thermometer**' (Duffy *et al.*, 2019; Gidron, Adams, & Horne, 2018; Iyengar *et al.*, 2012; Reiljan, 2020; Rogowski & Sutherland, 2016; Wagner, 2021). The former measure asks respondents to indicate their affect on a scale ranging from "dislike" to "like", and is for instance included in the Comparative Study of Electoral Systems (CSES) (Reiljan, 2020). The related feeling thermometer presents participants with a 0-to-100 point scale ranging from 'cold and negative' to 'warm and positive'. The American National Election Studies (ANES), which are often used by scholars study-

ing AP in the US, have long since included a thermometer scale to measure partisan affect (Iyengar *et al.*, 2012). While this long timespan brings large benefits, thermometers have some weaknesses. Individual differences are likely to play a big role in interpreting feeling thermometers, with some people having a warmer "baseline" than others (Wilcox, Sigelman, & Cook, 1989). Lastly, the translation of the thermometer question from the US, where Fahrenheit is commonly used, to Europe, where Celsius is the more familiar temperature scale, potentially influences results, but this has, to our knowledge, not yet been scrutinized in research.

Other scholars have intended to arrive at a measure of affect by analyzing **trust** in (supporters of) different parties (Druckman *et al.*, 2018; Druckman & Levendusky, 2019; Duffy *et al.*, 2019). This generally entails asking respondents to indicate on a scale how much they trust others. A more elaborate way to measure trust-levels, which goes beyond the measure of general attitudes or affect, is the use of "trust-games". Trust-games assess the extent to which participants are willing to donate or risk money they would otherwise receive themselves to co-partisans, while simultaneously withholding money from opposing partisans (Iyengar & Westwood, 2015). Research from the US and the UK has found that stereotypes, trust ratings, and feeling thermometers are strongly correlated (Druckman & Levendusky, 2019; Duffy *et al.*, 2019). Although trust measures and trust-games are both interesting strategies, it is important to remark that they fundamentally different in at least two aspects. First, trust scale capture an attitude, whilst trust games capture a behaviour. As psychological literature has long established, although (imperfectly) connected these two levels are conceptually distinct (Chaiklin, 2011). A second important aspect is that indeed trust-games do not solely capture trust, but also cooperation and civility, and they are used to investigate how trust is affected by different factors, such as social norms, culture, and cognitive reflection (Gong & Liu, 2021).

Yet another approach has been to ask respondents which **traits** describe the different parties and/or party-supporters (Almond & Verba, 1963; J. N. Druckman & Levendusky, 2019; Duffy *et al.*, 2019; Iyengar *et al.*, 2012). The traits respondents can choose from are both positive and negative and usually include attributes such as patriotic, closed-minded, intelligent, hypocritical, selfish, honest, open-minded, generous, and mean. Usually, scholars are not interested in the distinct content but rather the valence of these traits. An often-heard criticism to this measure is that it may be strongly biased by social desirability concerns. Respondents might hesitate to call someone selfish or unintelligent, which are quite

harsh judgements. A noteworthy alternative to circumvent social desirability concerns is presented by scholars employing a version of the Implicit Association test (IAT) in addition to directly asking respondents to rate their feelings about others (Iyengar & Westwood, 2015).

As a very extreme form of negative feelings, polarization research has recently started examining how individuals **dehumanize** members of their out-groups as a phenomenon connected to AP. According to Kteily and colleagues (2015), individuals' dehumanization of others is a natural consequence of the distinction between in- and out-groups. As AP induces in-group favouritism and out-group discrimination, it facilitates aggressive attitudes, intentions and even behaviours (Moore-Berg, Hameiri, & Bruneau, 2020). Multiple researchers have found partisans from both ends of the political spectrum to dehumanize the other (Martherus *et al.*, 2019; Moore-Berg *et al.*, 2020). Despite this close association between dehumanization and affective polarization, Martherus *et al.* (2019) argue that dehumanization is conceptually and empirically distinct from AP – or at least from the first facet of AP, general attitudes. To investigate this unique concept, scholars have used different measures of dehumanization, the more blatant being Kteily *et al.*'s (2015) visual dehumanization scale, which asks people to grade the humanity of others on a visual "ascent of man" scale.

Another category of AP-measures looks at **social distance** between people. This is also referred to as *social polarization*, *behavioral intentions* or the *level of intimacy* (Druckman & Levendusky, 2019; Duffy *et al.*, 2019). Rather than measuring attitudes, behavioral measures aim to determine the degree of AP based on how comfortable individuals are with forming intimate social bonds with members of their own and other parties. Hence, AP is high when respondents avoid social contact with individuals on basis of their political – partisan – identity and low if this is not the case (Duffy *et al.*, 2019). Different commonly used scenarios include individuals forming friendships (Duffy *et al.*, 2019; Levendusky & Malhotra, 2016), discussing politics (Duffy *et al.*, 2019; European Election Studies, n.d.), or having a son or daughter marrying someone from a certain party (Almond & Verba, 1963; Duffy *et al.*, 2019; Iyengar *et al.*, 2012). Klar *et al.* (2018) argue that the social distance measure conflates a dislike for out-party members with a dislike for partisanship, and show that oftentimes people simply seem to want to avoid talking about politics in general, regardless of political color of the conversation partner. Prior research also shows that, in the United States, indicators of general attitudes, like thermometer scales, have only a weak relationship with measures of social distance (Druckman & Levendusky, 2019; Duffy *et al.*, 2019), implying that, possibly, these measures capture different concepts.

AFFECTIVE POLARIZATION MEASUREMENTS IN CONTEXT

There are several reasons for which applying AP measures developed in the US context in European multiparty systems should not be considered as a completely unproblematic operation. The first one is methodological: respondents might express more gradual evaluations in contexts with more than two parties, which might lead to more divergence in item responses. The second reason, linked to the previous one, pertains to the fact that voters in a multiparty system are faced with not only multiple choices, but with different scenarios linked to these different choices (e.g., regarding coalition formation). In the US, elections are a zero-sum game. This is also demonstrated by the fact that those who do not like Democrats, are likely to be Republicans and vice versa, as in the US, 85-90% of voters feel close to or identify with one of these two parties (Petrocik, 2009).

But what about a multiparty system? There, even liking a party (to a certain extent) cannot be interpreted as being a steady supporter of that party. In contrast to the US, voters can easily switch from one party to another without necessarily crossing an ideological divide and might dislike a party for strategic reasons or based on some current coalition arrangement. This prominent difference between the US and European countries also shapes a different social context, in several respects. First, in a multiparty system the relationships among voters, party supporters and sympathizers are more nuanced and can be influenced by a variety of factors not only at the individual, but also at the systemic level (e.g., coalitions, signals among parties; Horne *et al.*, 2022). Second, in each country, there are different divides across which the preferences can be aligned: the traditional ideological left-right one, but also linguistic, territorial, cultural divides (see also Westwood *et al.*, 2017). And, as Hogg and Smith (2007) remark, the social context is a very important factor in shaping attitudes and identities because "group-defining attitudes are more likely to be reflected in behaviours when people identify strongly with a group" (p.120). Third, research has observed that in multi-party systems there are a number of other less stable and long-standing factors that can affect the relationship between voters and parties, such as issue preferences (Bartle & Bellucci, 2009), leader evaluations (Garzia, 2013) and past voting

behavior itself (Thomassen, 1976; Thomassen & Rosema, 2009). For all the reasons discussed, it seems clear that strong identification with a (social) group linked to a party is way more likely to happen in the US rather than in a European multiparty system.

In sum, we have reason to think that affective polarization could be influenced by the social and institutional context. With the rising scholarly interest that AP is currently enjoying in Europe, it becomes relevant to understand how AP measures perform one vis-à-vis the other in this quite different social and institutional setting. Our study sets out to test exactly that.

METHOD

Rationale

Research suggests that, when a construct is still unknown and not directly observable (as is the case for AP in Europe), the best strategy is to develop a multi-item instrument (see e.g., Fayers & Hand, 2002). However, this comes at the expenses of the length of the questionnaire, which is also a pivotal aspect. In order to understand whether, as Druckman & Levendusky (2019) found in US, some measures are comparable for European-based repondents as well, we developed a questionnaire including many of the aforementioned measures: (1) the feeling **thermometer** (Iyengar *et al.*, 2012)**;** (2) **like-dislike scores** (Wagner, 2021)**;** (3) **trust** (4) **dehumanization**; and (5) different levels of **social distance** (Levendusky & Malhotra, 2016).

Of course, including all these different operationalizations in the same study might create convergence in the answers, if only out of a consistency motivation. This means that we might overestimate similarities between answers. However, it is important to note that respondents filled out an entire battery (consisting of up to 9 parties, dependent on country) for one particular outcome variable before moving on to the next, which was presented on a new screen. This means that it would require quite some cognitive strain to remember all the exact answers provided on a previous screen.

Data collection

This survey was employed in a convenience sample of international students from nine different European nationalities at Maastricht University (UM) in the Netherlands in December 2020 and January 2021. According to the QS World University Ranking 2019, the student population of UM (about 18,000 students) is the 8th most international in Europe, with more than 50% of the students coming from other countries – a feature that serves particularly well in this case, as national background is a key element. Despite the obvious limitations due to the population composition (truncated demographics and high education), we consider the Maastricht University setting to be a suitable environment to test the functioning of several questions (e.g., the ones related to parties in each country) and to highlight potential pitfalls and country differences, both methodologically and substantially. The European nationalities most-represented in the student population of UM are Belgium, France, Germany, Greece, Spain, Italy, the Netherlands, Poland, and the United Kingdom. Therefore, respondents were only eligible for participation in the survey if they had one of these nationalities, as well as eligibility to vote in the country of nationality.

In our survey, of the 423 respondents who started, 327 completed 100% of the survey. 15 respondents were dropped because they did not fit nationality demands, were not eligible to vote, or did not pass the attention check question. This leaves us with a total of 312 respondents (115 male; 193 female; 5 non-binary) with a mean age of 22 years old (18 min; 43 max). 70.93% of respondents reported to be BA students, 25.56% were MA students, and 3.51% is following another type of education (e.g., just finished a degree, or doing a premaster). The distribution of the different nationalities is presented in Table 1.

Student samples started being widely used in explorative research in the 60s. The use of non-representative student samples has often been criticized especially because of their lack of generalizability potential (Benz & Meier, 2008; Brewer & Gros, 2010, p. 167; Cappella & Jamieson, 1997; Sears, 1986). Especially Sears (1986) expressed concerns with regard to differences between students and

Table 1. Percentage of respondents per nationality.

Nationality	% Respondents
Belgium	8.63
[Flemish]	[5.81]
[Walloon]	[2.45]
Germany	32.91
Greece	5.11
Italy	12.46
The Netherlands	21.73
Poland	6.71
Spain	2.56
France	7.67
The UK	2.24

non-students. However, recent research by Krupnikov *et al.* (2021) found that "much of the empirical research on the use of convenience samples suggests that the results obtained using these samples often replicate the results obtained with probability samples" (p. 179). All in all, as Cappella and Jamieson (1997) pointed out, the problem boils down to the fact that "students are different in education, ideology, political knowledge, experience and age from the voting public or the population as a whole" (Cappella & Jamieson, 1997). However, if for some characteristics such as age and political sophistication, the student sample cannot estimate an effect comparable to the general population, for others it is a viable choice. Aarøe (2011) conducted an experiment on a student sample of Danish university students and found the sample to be representative compared to the broad public with regards to important characteristics such as political interest, predispositions, and voting behaviour (Aarøe, 2011). Representativeness aside, note that our interest lies not in producing point estimates of some quantity in the population (say, the percentage of Belgians being affectively polarized), but rather in correlations between items. These are likely less impacted by the composition of the sample. Finally, using a student sample comes with two main advantages. First, we could assume a high level of education and administer a quite long and detailed survey with overall minor concerns about respondents' ability to focus for a long span of time (the questionnaire took about 30 minutes). Second, administrating such a long questionnaire in several countries would have been extremely expensive, which is additionally challenging given that the final objective was of a methodological nature. Although coming with the downside of not allowing very specific intra-country analyses due to the small number of respondents, the fact that several nationalities are represented in this data collection limits the risk that the results are country-specific.

Measurements

Affective polarization

Regarding the object of polarization, all AP questions were asked about party *supporters.* In addition, items were repeated for the prominent politicians of these parties in the case of the like dislike, trust, and the feeling thermometer batteries. A particular partisan group (supporters and prominent politicians) was included in the survey if they were represented in a country's national parliament at the time this survey was fielded. Moreover, for countries with a large number of parties in the national parliament, a (large) selection of the biggest and most extreme parties was included. For these decisions, experts on the different countries were consulted. The maximum number of parties included for one country in the survey is nine, which is the case for both Spain and the Netherlands. An overview over the parties selected for each country is attached in Appendix A.

To measure affective polarization, we included different ways of asking respondents about their general attitudes and feelings. First, respondents were asked to use a 0-10 scale to respectively indicate their degree of **dislike** (0) or **like** (10) towards both voters and leading politicians. Measuring like-dislike for both voters and elites can give researchers insights into whether there is a difference between the so so-called vertical and horizontal dimensions of affective polarization, were the first one pertains to polarization towards the elites, and the second to the one towards fellow citizens (see Harteveld, 2021). However, as we found a high correlation (0.835) between the two measures (as many before us – see for instance Druckman & Levendusky, 2019; Harteveld 2021), we decided to focus on the horizontal dimension of affective polarization, which is the one originally conceived by Iyengar and Westwood (2015) ("view[ing] opposing partisans negatively and copartisans positively" – p. 691) and Huddy and Yair (2020) ("hostility between rival political partisans" – p. 1).[1] We then continued to ask them about their **trust** (0, or not at all, to 10, completely) towards both types of objects. Evaluating trust in political actors, such as politicians or government institutions, can shed light on how perceptions of trustworthiness influence voting decisions. High levels of trust may lead to increased support, while distrust can result in opposition. Subsequently, respondents filled out a **thermometer scale** (0 cold-100 warm) for those groups. This provides a quantitative measure of emotional responses, which can be used to understand how emotional affinity or hostility affects voting choices. To measure **dehumanization**, we used Kteily *et al.*'s (2015) measure of dehumanization and asked respondents to place the voters of the different parties in their country on a scale using the "ascent of man" picture (see Appendix B). Understanding the extent of dehumanization can reveal the impact of negative campaigning on voter attitudes and behavior. If voters perceive opponents as less than human, it can lead to more hostile and divisive political environments. To assess **social distance**, we included questions on how comfortable or uncomfortable respondents would be in different social relationships with voters of certain parties (0-10 scale). Social distance questions give us a direct indication of the so-called horizontal polarization in a given context.

---

[1] For examples of studies focusing on affective evaluation of leaders see Barisione (2017) and Bordignon (2020).

All these measures, especially when compared from context to context can help us understanding the extent to which countries encounter similar dynamics. Questions were asked about relations with different degrees of closeness, namely having a romantic relationship; being close friends; being loose acquaintances; having a close friend being in a romantic relationship with someone. In contrast to the like-dislike, trust and thermometer questions, the dehumanization question and the social distance question were not asked for political leaders, only for party supporters. This decision was based partly on the practical concern of an overly lengthy survey, and partly on the fact that social scenarios involving leading politicians may not be very realistic.

Some prominent measures were not included in our questionnaire, among which party feeling thermometer scale and a traits battery. The main reason for these exclusions was practical. The questionnaire was already very long compared to current recommendations for online surveys, and both these questions require a substantial amount of additional time to be answered. Furthermore, both these measures were found to be highly correlated with the voters feeling thermometer (Iyengar *et al.*, 2012; Druckman & Levendusky, 2019; Gidron *et al.*, 2022). Finally, aggregate measures based on traits are multi-item measures that are not so easily compared with our other scales.

Connected concepts

To map respondents' political identity, **partisanship** was measured by asking respondents what party they feel closest and what party they feel most distant to. (Almond & Verba, 1963; Iyengar *et al.*, 2012; Reiljan, 2020; Wagner, 2021). Three questions were asked to measure respondents' **political interest**. We asked respondents how interested they are in politics, how closely they follow what goes on in government and politics, and how often they discuss politics and current affairs with others (ESS, 2018).

Procedure

The survey was administered in Qualtrics and disseminated through student Facebook-groups and in-class promotion. Participation was voluntary.. Participation was incentivized: at the end of the survey participants could leave their email address to partake in a lottery in which one gift voucher of 100 euros and four gift vouchers of 50 euros were allotted and were thanked for their participation. Until twenty years ago there was a quite broad consensus about the fact that lottery incentives did not significantly impact survey participation (Church, 1993; Singer, Hoewyk, & Maher, 2000; Warriner *et al.*, 1996). However, it has since been shown by using web-based surveys with student samples that lottery incentives increase both participation and completion rate (Bosnjak & Tuten, 2003, p. 215; Cobanoglu & Cobanoglu, 2003, p. 485; Laguilles, Williams, & Saunders, 2011, p. 549; Porter & Whitcomb, 2003, p. 403). For instance, Porter and Whitcomb (2003) found that the amount of the incentive decreases after a certain. They experimented with different amount of money ranging from $50 to $200, and found that the marginal effect of participating decreased substantively after $50.

Firstly, respondents read a short introduction and were asked for their informed consent. In the introduction, we asked respondents to act as a political expert on their country of origin and informed that they would only be eligible for participation in the survey if they hold citizenship in Belgium, France, Germany, Greece, Spain, Italy, the Netherlands, Poland, or the United Kingdom. They were furthermore warned that the survey could be repetitive and asked to still answer each question carefully. We informed the students that the survey was a long one, as previous literature found that it is important to make respondents aware of the duration in advance in order to minimize dropping (Galesic & Bosnjak, 2009; Hansen, 2007).

After the intro, respondents answered filtering questions on their nationality and eligibility to vote. Eligible respondents answered AP-questions on general attitudes, social distance, and dehumanization. After this, respondents saw an attention check for which they had to move the slider in the question all the way to the right. Next, respondents answered the different questions on partisan identity and political interest and lastly some questions on socio-demographic characteristics.

Analysis

In order to properly investigate and contrast the different AP measures in the different countries, the dataset was reshaped and stacked, to arrive at a triadic data structure. This means that for each respondent, the dataset contains as many observations as there are parties in their country times five (the amount of AP measures used). Hence, there is an AP score for every respondent-party-measure combination. For a respondent from France, for instance, the dataset would contain 35 observations, as seven parties are included for France and five AP measures are used. The advantage of this setup is that is allows to predict answer patterns by features of the respondent, measure, and party simultaneously.

The thermometer and dehumanization scores were recoded to a 0-10 scale, to match the other measurements, and, to prevent respondents from simply repeating their answers, scales were occasionally reversed so that higher scores indicate more negative evaluations.

Our analysis proceeds as follows. First, we aim to establish whether answer patterns differ systematically between measures, countries, and targets. We do so through a descriptive analysis (step 1) and a formal test in a multivariate model (step 2). After doing so, we proceed to assess if the items reflect a single construct or multiple constructs. For this, we use explanatory and confirmatory factor analyses (step 3) and predict the different subsets that come out of it (step 4).

RESULTS

Different measures, different answers?

<u>Step 1: Descriptives</u>
Before moving to our main analyses, Figure 1 below presents the mean scores for all measures, for three types of parties: the party the respondent indicated they feel closest to (the partisan question); the one they feel furthest from; and all others ('not closest, not furthest'). Figure 2, in addition, shows the distribution of scores on the different measures for respondents' most and least liked group of other voters. As noted, all variables were rescaled to 0 to 10, with higher scores indicating more negative evaluations.

Clearly, all measures pick up on a difference in evaluation between the respondents' closest and furthest (as well as all other) parties. Note that these differences are very substantial: up to 8 points on the 11-point scale. Importantly, the items 'liking', 'trusting', or 'having warm feelings towards' a political outgroup all yield quite similar scores. This suggests that the actual wording of the scale extremities is not crucial, as long as they refer to some form of affective evaluation. By contrast, the social distance scores differ in their point estimates from the first three items as well as between themselves: envisaging a romantic engagement with an outgroup member yields similar average scores as the first three measures, whereas imagining a close acquaintance from the outgroup does not trigger such a negative response. While this is not surprising, given that the different items are developed to reflect different levels of intimacy and hence to differentiate 'easier' from 'harder' items, it is still important to note that some yield a nominal
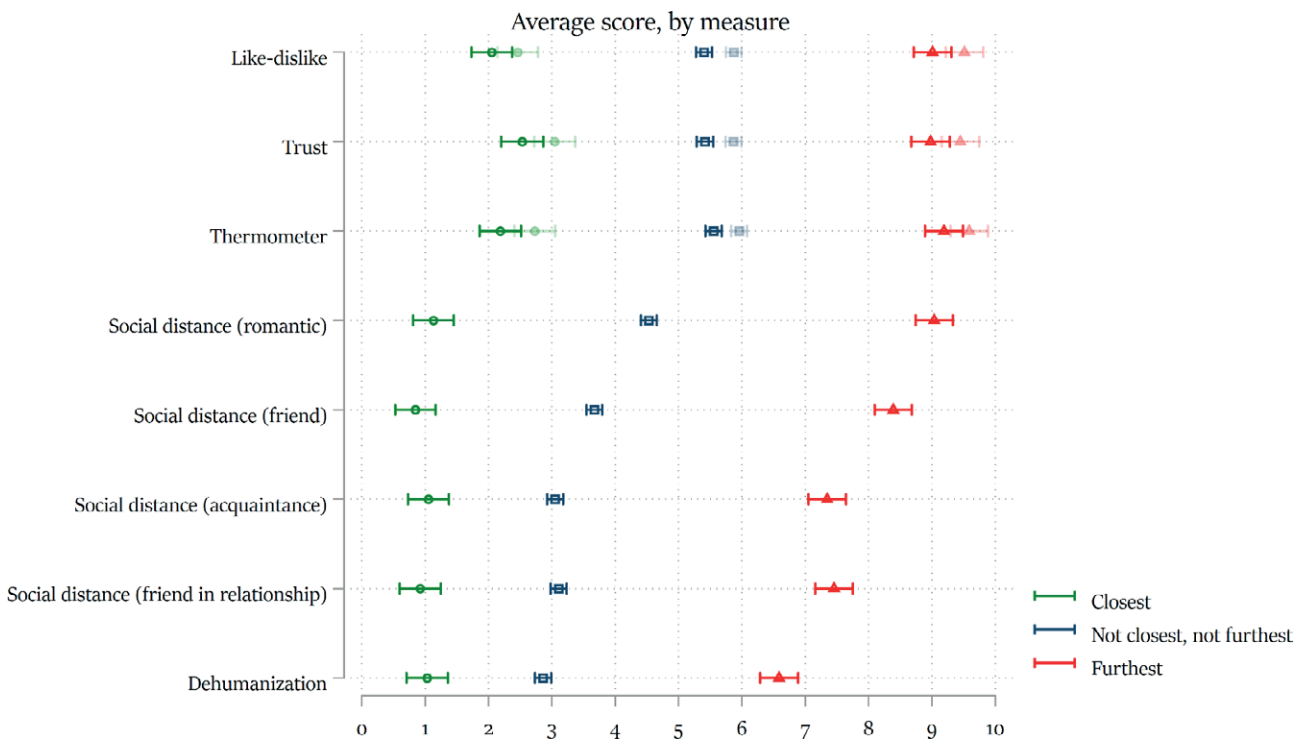
Figure 1. Average scores of the different measures. Note: higher scores denote more negative evaluations; all items rescaled 0-10; faint markers denote evaluations of politicians. With 95% confidence intervals.
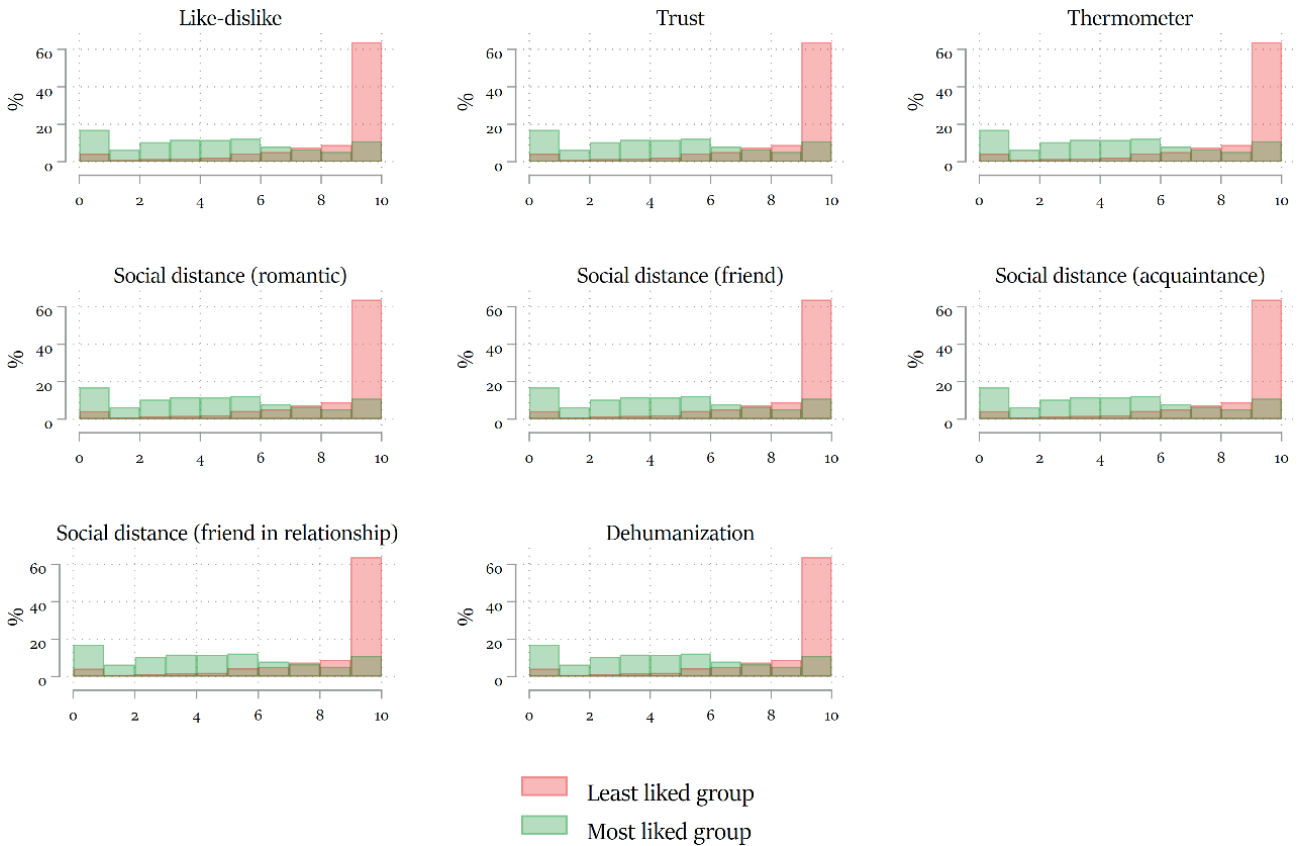
Figure 2. Distribution of scores for the different measures. Note: all items rescaled 0-10; based on scores toward voter groups.

distribution similar to the affective scales and some to the dehumanization measures. We can also notice that for the three attitudinal measures (like-dislike, trust scale, and thermometer), the scores stay consistently higher. They start to decrease with the behaviour measures (social distance), and in function of the distance the respondent has with the social object (in this case a person). They then further drop with another attitudinal question (dehumanization), but a very extreme one.

Although respondents from different country contexts provide different mean scores, in general, the patterns between the items are similar. In other words, all items deliver the same impression of the level of affective polarization in a context, and there is no clear evidence that some items yield very context-specific answers. For an overview of the average scores on the different measures by country, see Appendix C.

Step 2: Modelling the answers
To put the patterns suggested by Figures 1 and 2 to a formal test, Figure 3 below presents a regression model predicting respondents' score in the triadic data by characteristics of the measure, party, and individual.

Put differently, we regress the variable containing all AP scores a respondent has given (for all parties and all measures) on the different types of measures, the different countries, the targets and the relation to the party. Again, these analyses confirm that asking a 'like-dislike', trust, or thermometer scale yields no significant difference. Social distance questions generally produce evaluations that are up to 2 points more positive. The scores handed out by respondents socialized in different political systems also differ markedly, with Greek respondents showing most negativity and Walloon respondents least. Items targeting politicians rather than voters get somewhat more negative scores too, but only to a limited extent.

One construct or many?

Step 3: Factor analysis
The similarities in some of the answer patterns – especially between like, trust, and thermometers, as well as between some of the social distance items – beg the question whether the various measures tap into the same
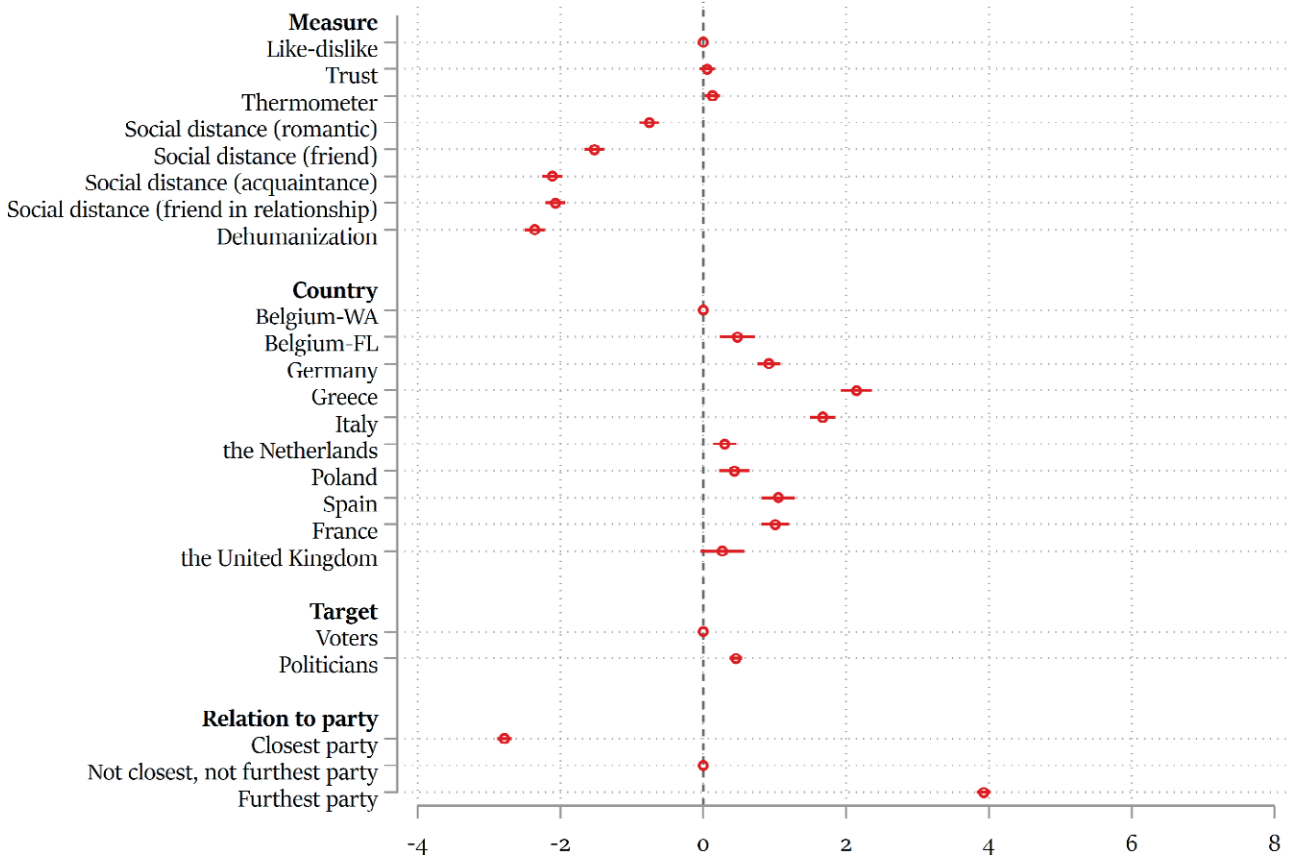
Figure 3. AP scores predicted by measure, country, target, relation to party. Note: Coefficients with 95% confidence bars. Reference categories are the open circles on the 0-line without confidence bars. Higher scores denote more negative evaluations.

construct or separate ones. This section therefore contains the result of an explanatory factor analysis (EFA) and confirmatory factor analysis (CFA). Because the party under evaluation will often matter more than the measure used, we restricted this factor analysis to the party the respondent feels furthest from.

An EFA of the eight different measures shows a strong Eigenvalue of over 4 for the first factor, and an Eigenvalue of 0.79 for the second. An investigation of the two-factor structure (Table 2) suggests that *like*, *trust*, and *therm* share a factor with the other items (with strong loadings except for dehumanization), and, in addition, a weak of their own (which is not shared by the others). Still, Table 2 strongly suggests that all items tap into a shared underlying factor to a very substantial degree.

To obtain a formal test, we proceeded by estimating different CFA models in turn, reported in Table 3. In the first model, all items were modelled to follow from a single latent construct. This model does not fit the data very well compared to the usual cut-off points of 0.05 for good and 0.08 for acceptable fit. An investigation of modification indices (MIs) suggests that the most important

sources of misfit are strong residual correlations between *like-dislike*, *trust*, and *thermometer*. In a second model, we loaded those on a separate construct. This improves the model fit but still not to a satisfactory degree. The MIs suggest one source of misfit is residual correlation between the items *social distance romantic* and *social distance friend*, as well as *social distance acquaintance* and *social distance friend in a relationship* (the two more distant relations). Providing these with separate constructs leaves us with dehumanization, which as a single item cannot be loaded on its own latent construct. We therefore leave it out of the model third model. At RMSEA=0.086 this one starts to reach an acceptable fit. The remaining MIs suggest that *trust* also loads on the dimension of the 'intimate social distance'. However, for theoretical reasons (the separation of affective responses and social distance intentions) we consider that it is most fruitful to think of the eight items to span four different but highly correlated clusters: *affective scales* (like, trust, and thermometer), *intimate social distance* (*romantic* and *friend*), *non-intimate social distance* (*acquaintance* and *friend in a relationship*), and *dehumanization*.

Table 2. Explanatory factor analyses.

Variable	Factor 1	Factor 2	Uniqueness	Variance
Like/dislike	0,75	0,34	0,33	0,67
Trust	0,56	0,48	0,46	0,54
Thermometer	0,72	0,38	0,34	0,66
Social distance (romantic)	0,77	-0,15	0,39	0,61
Social distance (friend)	0,82	-0,30	0,25	0,75
Social distance (acquaintance)	0,84	-0,23	0,25	0,75
Social distance (friend in relationship)	0,78	-0,36	0,26	0,74
Dehumanization	0,49	0,06	0,76	0,24

Table 3. Confirmatory factor analysis.

Model	Description	RMSEA
1	All eight items load on one latent construct	0.222
2	Separate latent constructs for (1) *like*, *trust* and *therm* and (2) all others	0.128
3	Separate latent constructs for (1) *like, trust and therm*, (2) *social distance romantic* and *friend*; (3) *acquaintance* and *friend in relationship*. Drop *dehuman*	0.086

How do these various (sets of) indicators relate to each other? Table 4 shows that there is a very strong correlation (0.88) between the two types of social distance latent constructs, and a moderately strong one between the social distance latent constructs and the affective scales (0.62-0.64). The observed item of dehumanization correlates only weakly (<0.47) with either of those. In short, although differences exist between the two suggested social distance constructs, for practical purposes it is reasonable to collapse them and to distinguish between the categories of affective scales, social distance, and dehumanization.

Step 4: Who differentiates?
The analysis above suggests that a distinction can be made between three categories of items, but at the same time shows correlations between those categories to be strong – to the point that the EFA does not pick up on their differences. Still, it might be that this differentiation appears weaker because not all respondents make the distinction between measures and between targets (voters vs politicians). In particular, it is likely that politically interested individuals do so more clearly. If this were to be the case, then it might still pay off to use multiple items to study the subgroup of the politically interested.

To test this expectation, we interacted the different measures, as well as in the case of the affective scales the different targets with political interest in the triadic dataset. This test shows that among the politically interested the answer patterns for the dehumanization item differ significantly from the affective scales, compared to respondents scoring lower on political interest. No interaction was found between political interest and the target. Figure 4 visualizes these models. It shows clear main effects: the more politically interested, the more negative respondents are, in line with the literature. Furthermore, affective scales and politicians evoke more negative scores than the other measures and voters. However, the dehumanization items do not depend on political interest as much as the others. As a consequence, politically interested voters make stronger distinctions between affective measures and social distance on the one hand and dehumanization on the other. However, the interaction is not very substantive, and given that most scholars' interested will lie with less extreme forms of political outgroup bias, we conclude that answer patterns are relatively similar regardless of political sophistication.

CONCLUSIONS

The literature on affective polarization is thriving, and important strides have been made to conceptualize and operationalize this concept. Affective polarization measures are valuable tools for comparative research into the determinants of voting behavior. They provide a deeper understanding of the emotional and attitudinal aspects of politics, helping researchers navigate the com-

Table 4. Correlation coefficients between the three latent scales and dehumanization item.

	Affective scales (like, trust, thermometer)	Intimate social distance	Non-intimate social distance	Dehumanization
Affective scales (like, trust, thermometer)	1.00			
Intimate social distance	0.64	1.00		
Non-intimate social distance	0.62	0.88	1.00	
Dehumanization	0.44	0.39	0.47	1.00

Political interest
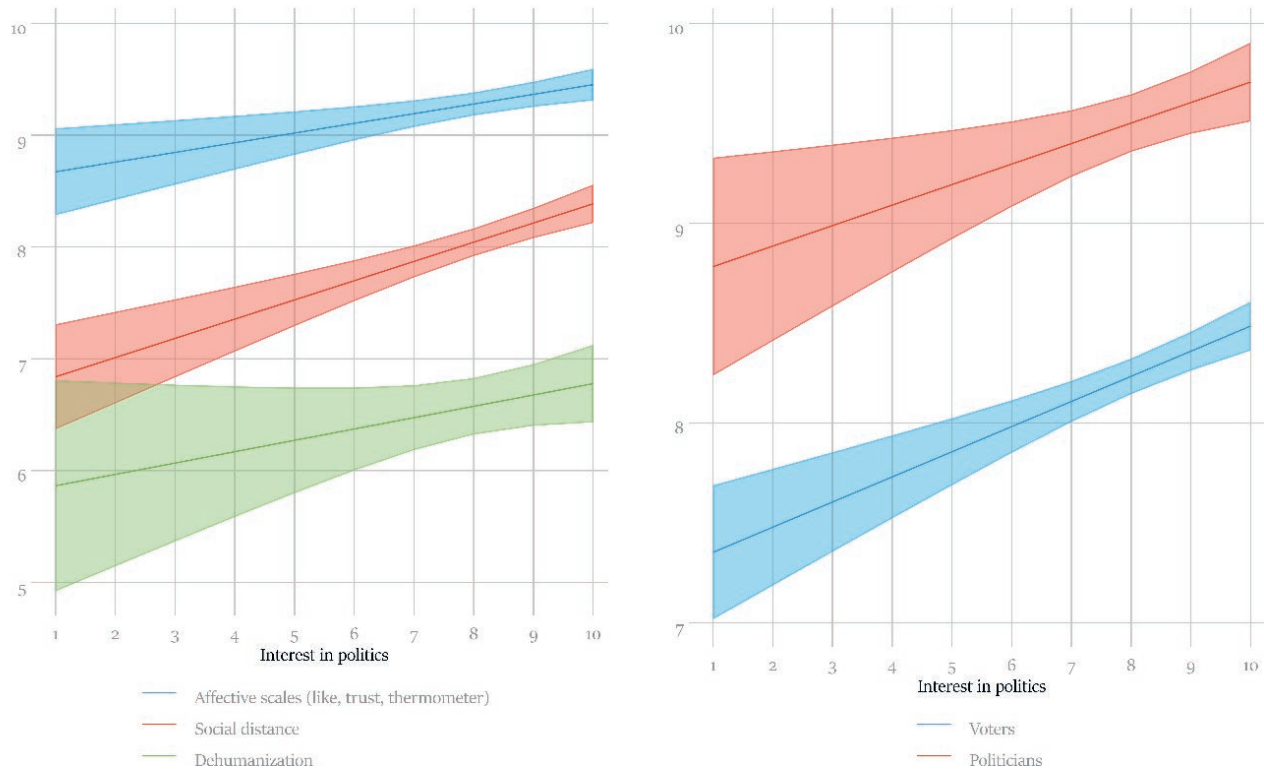
Figure 4. Interaction with political interest.

plexities of electoral dynamics and political behavior in various contexts. However, deciding which operationalization to employ among the several that are available. The aim of our paper is to understand how all these measures relate to one another. The choice of which operationalization to employ needs to be informed by theoretical choices, the research question, and the overall design of the research. Yet, with an increasing number of new original data collections, having some indication of these constructs performed in comparison to one another becomes pivotal in making an informed choice.

In this paper we have used a sample spanning students with nine different European national backgrounds to investigate whether different types of measures reflect different concepts or whether they are simply different variations of the same measurement. In a seminal contribution, Druckman and Levendusky (2019) conducted a similar test on an US-based sample. However, the differences in the setting and dynamics between the US and European countries – including the diverging role of political identities, the more gradual evaluations that are possible in multiparty contexts, and different meanings attributed to the words and behaviours mentioned in item wordings – call for an empirical test. Our aim is to

contribute to the understanding of affective polarization and its measurement by assessing how citizens socialized in different multiparty contexts interpret and use these different types of measurements. Our analysis allows for the formulation of a number of conclusions.

First, all types of items produce strong differences between respondents' 'in-party' and 'out-party', up to 8 points on the 11-point scale. In a way, any item involving an evaluation of the out-party will produce highly differentiated answers that differ much more between parties than between items. In other words, studies aiming to capture affective polarization with a broad brush might be relatively free to use any of the instruments suggested in the literature.

Second, it is striking that – for all practical purposes – respondents did not differentiate in their '[dis]like', '[dis]trust' or 'warm [cold] feelings' towards political outgroups. This is noteworthy because these operationalizations stem from different traditions in the study of political behavior and are often argued to capture different phenomena – for instance, thermometers being more 'affective' than 'dislike'. Of course, the design of the study (which involves within-person comparisons of batteries) is likely to create a convergence of answers.

Still, it is noteworthy that respondents did provide quite dissimilar answers on some of the other scales. Hence, it seems justified to conclude that items that involve some affective evaluation of outgroups (with positively and negatively valanced terms on the extreme) will produce very similar point estimates.

Third, the various social distance items produced different point estimates (as expected, depending on intimacy). They also produce slightly different response patterns, but, for practical purposes, can be usefully combined into a single indicator, which in turn correlates moderately strongly (around .63) with the affective scales. This is in line with previous literature, which tends to move towards approaching social distance as different from, although related to, affective polarization proper (Klar *et al.*, 2018). Dehumanization stands out as a very different phenomenon, correlating only weakly with the others, and having somewhat different predictors.

Fourth, also in line with previous literature (Druckman & Levendusky, 2019), politicians receive lower sympathy than voters. Using items based on abstract entities or even explicitly mentioning politicians will therefore yield higher observed levels of affective polarization than items describing average voters. Still, the correlates of both types appear roughly similar, which suggests items about politicians can be used with some caveats to study the antecedents of affective polarization as a horizontal phenomenon.

Fifth, we found little reason to worry that the items operate very differently across different contexts. Admittedly, our student sample is not representative and still relatively homogeneous in terms of political socialization. Still, it is telling that, while we found strong differences in mean scores – students with a Greek nationality providing scores that are more than 2 points more negative than the least polarized group, the Belgians – we found little evidence that response patterns to individual items differed between countries.

All in all, these results bears good news to the existing and future practice of operationalizing affective polarization. The choice of scale appears less influential than might be expected given the relevant differences between the US and Europe as well as between European national contexts. This is especially true when using items that contain a negatively and positively valanced endpoint of the scale. This suggests that scholars studying affective polarization in multiparty systems can rely on a single battery (one item for reach party), which strongly reduces the survey space needed to measure this concept in fragmented political landscapes. Still, from a theoretical point of view, there seem to be several reasons suggesting to employ the largely used like-dislike scale only when other, more precise and definite operationalizations do not fit the purpose of the research. As discussed, in multiparty systems, the reasons and the implications of (dis)liking a party are broader than in a two-party system, making inferences about negative affect less clearcut. Future research could explore how the measurements under study operate in diverse contexts by assessing what citizens have in mind when evaluating political outgroups (Druckman *et al.*, 2022).

REFERENCES

Aarøe, L. (2011). Investigating Frame Strength: The Case of Episodic and Thematic Frames. *Political Communication*, *29*(2), 207–226.

Achterberg, P., & Houtman, D. (2009). Ideologically illogical? Why do the lower-educated dutch display so little value coherence? *Social Forces*, *87*(3), 1649–1670. https://doi.org/10.1353/sof.0.0164

Almond, G. A., & Verba, S. (1963). *The Civic Culture – Political Attitudes and Democracy in Five Nations*.

Arian, A., & Shamir, M. (1983). The primrily political functions of the left-right continuum. *Comparative Politics*, *15*(2), 139–158.

Barisione, M. (2017). The Partisan Gap in Leader Support and Attitude Polarization in a Campaign Environment: The Cases of Germany and Italy. I*nternational Journal of Public Opinion Research*, 29 (4): 604–30.

Bartle, P., & Bellucci, J. (Eds.). (2014). P*olitical Parties and Partisanship: Social identity and individual attitudes.* Routledge.

Benz, M., & Meier, S. (2008). Do People Behave in Experiments as in the Field?: Evidence from Donations. *Experimental Economics*, *11*, 268–81.

Bordignon, F. (2020). Leader Polarisation: Conflict and Change in the Italian Political System. *South European Society and Politics*, *25*(3/4), 1–31.

Bosnjak, M., & Tuten, T. L. (2003). Prepaid and Promosied Incentives in Web Surveys – an experiment. *Social Science Computer Review*, *21*(2), 208–2016.

Brewer, P. R., & Gros, K. (2010). Studying the Effects of Framing on Public Opinion about Policy Issues. In D'Angelo & J. A. R. Kuypers (Eds.), *Doing News*

Frame Analysis Empircal and Theoretical Perspectives. New York.

Cappella, J. N., & Jamieson, K. H. (1997). *Spiral of cynicism: The press and the public good*. New York: Oxford University Press.

Caprara, G. V., & Vecchione, M. (2018). On the Left and Right Ideological Divide: Historical Accounts and Contemporary Perspectives. *Political Psychology*, *39*, 49–83.

Caughey, D., O'Grady, T., & Warshaw, C. (2019). Policy Ideology in European Mass Publics, 1987 – 2016. *American Political Science Review*, 1–20.

Chaiklin, H. (2011). Attitudes, Behavior, and Social Practice. *The Journal of Sociology & Social Welfare*, *38*(1), 31-54.

Church, A. (1993). Estimating the effect of incentives on mail survey response rates: a meta-analysis. *Public Opinion Quarterly*, *57*, 62–79.

Cobanoglu, C., & Cobanoglu, N. (2003). The effect of incentives in web surveys: application and ethical considerations. *International Journal of Market Research*, *45*(4), 475–488.

Dalton, R. J. (2008). The Polarization of the European Party Systems – New Data , New Approach , New Results. *Comparative Political Studies*, *41*(7).

Druckman, J. N., Gubitz, S. R., Levendusky, M. S., & Lloyd, A. M. (2018). How Incivility On Partisan Media (De-)Polarizes the Electorate. In *American Political Science Association*.

Druckman, J. N., & Levendusky, M. S. (2019). What Do We Measure When We Measure Affective Polarization? *Public Opinion Quarterly*, *83*(1), 114–122.

Druckman, J. N., Klar, S., Krupnikov, Y., Levendusky, M., & Ryan, J. B. (2022). (Mis) estimating affective polarization. *The Journal of Politics*, *84*(2), 1106-1117.

Duffy, B., Hewlet, K., McCrae, J., & Hall, J. (2019). *Divided Britain? Polarisation and fragmentation trends in the UK*. Sovremennaya Evropa. https://doi.org/10.2307/622618

ESS. (2018). European Social Survey. ESS 9 Source Questionnaires.

European Election Studies. (n.d.). The European Election Studies.

Fayers, P. M., & Hand, D. J. (2002). Causal variables, indicator variables and measurement scales: An example from quality of life. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, *165*(2), 233–253. https://doi.org/10.1111/1467-985X.02020

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly*, *73*(2), 349–360.

Garrett, R. K., Dvir Gvirsman, S., Johnson, B. K., Tsfati, Y., Neo, R., & Dal, A. (2014). Human Communication Research Implications of Pro-and Counterattitudinal Information Exposure for Affective Polarization. *Human communication research*, *40*(3), 309-332.

Garzia, D. (2013). The Rise of Party/Leader Identification in Western Europe. P*olitical Research Quarterly*, 66(3), 533–544.

Gong, Z., Tang ,Y., & LIU C. (2021) Can trust game measure trust?. *Advances in Psychological Science*: *29*(1), 19-30.

Gidron, N., Adams, J., & Horne, W. (2018, August). How ideology, economics and institutions shape affective polarization in democratic polities. In Annual conference of the American political science association.

Gidron, N., Sheffer, L., & Mor, G. (2022). Validating the feeling thermometer as a measure of partisan affect in multi-party systems. *Electoral Studies*, *80*, 102542.

Hansen, K. M. (2007). The Effects of Incentives, Interview Length, and Interviewer Characteristics on Response Rates in a CATI-Study. *International Journal of Public Opinion Research*, *19*(1), 112–121.

Harteveld, E. (2021). Fragmented foes : affective polarization in the multiparty context of the Netherlands. *Electoral Studies*, *71*, 1–22.

Harteveld, E., Mendoza, P., & Rooduijn, M. (2021). Affective Polarization and the Populist Radical right -Creating the Hating. *Government and Opposition*, 1–25.

Hawkins, K. A., Carlin, R. E., Littvay, L., & Kaltwasser, C. R. (2018). *The Ideational Approach to Populism: Concept, Theory and Analysis*. Routledge.

Hogg, M. A., & Smith, J. R. (2007). Attitudes in social context: A social identity perspective. *European Review of Social Psychology*, *18*(1).

Hooghe, L., Marks, G., & Wilson, C. J. (2002). Does left/right structure party positions on European integration? *Comparative Political Studies*, *35*(8), 965–989.

Iyengar, S., & Krupenkin, M. (2018). The Strengthening of Partisan Affect. *Political Psychology*, *39*, 201–218.

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, September). Affect, not ideology: A social identity perspective on polarization. *Public Opinion Quarterly*.

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.

Kekkonen, A., & Ylä-Anttila, T. (2021). Affective blocs: Understanding affective polarization in multiparty systems. *Electoral Studies*, *72*.

Kingzette, J. (2021). Who Do You Loathe? Feelings toward Politicians vs. Ordinary People in the Opposing Party. *Journal of Experimental Political Science*, *8*, 75–84.

Klar, S., Krupnikov, Y., & Ryan, J. B. (2018). Affective polarization or partisan disdain? Untangling a dislike for the opposing party from a dislike of partisanship. *Public Opinion Quarterly*, *82*(2).

Kleiner, T.-M. (2018). Public opinion polarisation and protest behaviour. *European Journal of Political Research*, *57*, 941–962.

Krupnikov, Y., Nam, H. H., & Style, H. (2021). Convenience Samples in Political Science Experiments. In J. Druckman & D. P. Green (Eds.), *Advances in Experimental Political Science* (pp. 165–183). Cambridge: Cambridge University Press.

Kteily, N., Bruneau, E., Waytz, A., & Cotterill, S. (2015). The ascent of man: Theoretical and empirical evidence for blatant dehumanization. *Journal of Personality and Social Psychology*, *109*(5), 901–931.

Laguilles, J. S., Williams, E. A., & Saunders, D. B. (2011). Can Lottery Incentives Boost Web Survey Response Rates? Findings from Four Experiments. *Research in Higher Education*, *52*(5), 537–553.

LeBas, A. (2018). Can Polarization Be Positive? Conflict and Institutional Development in Africa. *American Behavioral Scientist*, *62*(1), 59–74.

Lelkes, Y. (2016). The polls-review: Mass polarization: Manifestations and measurements. *Public Opinion Quarterly*. Oxford University Press.

Levendusky, M., & Malhotra, N. (2016). Does Media Coverage of Partisan Polarization Affect Political Attitudes? *Political Communication*, *33*(2), 283–301.

Levendusky, M. S. (2010). Clearer cues, more consistent voters: A benefit of elite polarization. *Political Behavior*, *32*(1), 111–131. https://doi.org/10.1007/s11109-009-9094-0

Luskin, R. C. (1987). Measuring Political Sophistication. *American Journal of Political Science*, *31*(4), 856–899.

Martherus, J. L., Martinez, A. G., Piff, P. K., & Theodoridis, A. G. (2019). Party Animals? Extreme Partisan Polarization and Dehumanization. *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. (2018a). Ideologues without issues: The polarizing consequences of ideological identities. *Public Opinion Quarterly*, *82*(1), 866-887.

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

McCoy, J., Rahman, T., & Somer, M. (2018). Polarization and the Global Crisis of Democracy: Common Patterns, Dynamics, and Pernicious Consequences for Democratic Polities. *American Behavioral Scientist*, *62*(1).

Merkley, E., Cutler, F., Quirk, P. J., & Nyblade, B. (2019). Having their say: Authority, voice, and satisfaction with democracy. *Journal of Politics*, *81*(3), 848–861.

Moore-Berg, S. L., Hameiri, B., & Bruneau, E. (2020). The prime psychological suspects of toxic political polarization. *Current Opinion in Behavioral Sciences*, *34*, 199–204.

Munzert, S., & Selb, P. (2017). Measuring Political Knowledge in Web-Based Surveys: An Experimental Validation of Visual Versus Verbal Instruments. *Social Science Computer Review*, *35*(2), 167–183.

Ondercin, H. L., & Lizotte, M. K. (2020). You've Lost That Loving Feeling: How Gender Shapes Affective Polarization. *American Politics Research*, *49*(3), 282–292.

Oosterwaal, A., & Torenvlied, R. (2010). Politics Divided from Society? Three Explanations for Trends in Societal and Political Polarisation in the Netherlands. *West European Politics*, *33*(2), 258-279.

Petrocik, J. R. (2009). Measuring party support: Leaners are not independents. *Electoral Studies*, *28*(4).

Poole, K. T. (2011). The Roots of the Polarization of Modern US Politics. *SSRN Electronic Journal*, (September 2008).

Porter, S. R., & Whitcomb, M. E. (2003). The impact of lottery incentives on student survey response rates. *Research in Higher Education*, *44*(4), 389–407.

Ran, W., Yamamoto, M., & Xu, S. (2016). Media multi-tasking during political news consumption – a relationship with factual and subjective political knowledge. *Computers in Human Behavior*, *56*, 352–359.

Rapeli, L. (2014). *The Conception of Citizen Knowledge in Practices of Democracy*. Palgrave Pivot.

Reiljan, A. (2020). 'Fear and loathing across party lines' (also) in Europe: Affective polarisation in European party systems. *European Journal of Political Research*, *59*(2), 376–396.

Rogowski, J. C., & Sutherland, J. L. (2016). How Ideology Fuels Affective Polarization. *Political Behavior*, *38*, 485–508.

Russo, L., Franklin, M. N., & Beyens, S. (2021). Clarity of voter choices: neglected foundation for ideological congruence. *Quaderni Dell'Osservatorio Elettorale – Italian Journal of Electoral Studies*, *83*(2), 3–13.

Sears, D. O. (1986). College Sophomores in the Laboratory: Influence of a Narrow Data Base on Social Psy-

chology's View of Human Nature. *Journal of Personality and Social Psychology*, *51*, 515–30.

Singer, E., Hoewyk, J., & Maher, M. (2000). Experiments with incentives in telephone surveys. *Public Opinion Quarterly*, *64*, 171–188.

Sociaal en Cultureel Planbureau. (2016). Thematische collectie: Culturele Veranderingen in Nederland (CV) en SCP Leefsituatie Index (SLI). *Centraal Bureau Voor de Statistiek*.

Spoon, J. J., & Klüver, H. (2015). Voter polarisation and party responsiveness: Why parties emphasise divided issues, but remain silent on unified issues. *European Journal of Political Research*, *54*, 343–362.

Strabac, Z., & Aalberg, T. (2011). Measuring political knowledge in telephone and web surveys: A cross-national comparison. *Social Science Computer Review*, *29*(2), 175–192.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*(2).

Thomassen, J. J. A. (1976). P*arty identification as a cross-national concept: Its meaning in the Netherlands*. In: Party identification and beyond: Representations of voting and party competition (pp. 63–79). Wiley.

Thomassen, J. J. A., & Rosema, M. (2009). *Party identification revisited*. In: Political Parties and Partisanship. Social identity and individual attitudes (pp. 42–59). Routledge. https://research.utwente.nl/en/publications/party-identification-revisited

Wagner, M. (2021). Affective polarization in multiparty systems. *Electoral Studies*, *69*, 102199.

Warriner, K., Goyder, J., Gjertsen, H., Hohner, P., & McSpurren, K. (1996). Charities, no; Lotteries, no; Cash, yes. Main effects and interactions in a Canadian incentives experiment. *Public Opinion Quarterly*, *60*, 542–562.

Webster, S. W., & Abramowitz, A. I. (2017). The Ideological Foundations of Affective Polarization in the US Electorate. *American Politics Research*, *45*(4).

Westwood, S. J., Iyengar, S., Walgrave, S., Leonisio, R., Miller, L., & Strijbis, O. (2018). The tie that divides: Cross-national evidence of the primacy of partyism. *European Journal of Political Research*, *57*(2), 333–354.

Wilcox, C., Sigelman, L., & Cook, E. (1989). Some like it hot. *Public Opinion Quarterly*, *53*(2), 246–257.