How to Measure Ideological Polarization in Party Systems

Author: Johannes Schmitt, M.A. (Contact: Johannes.Schmitt@hhu.de)
Social Sciences Institute - Comparative Politics, Heinrich Heine University (Düsseldorf)

Paper prepared for the ECPR Graduate Student Conference 2016 (University of Tartu)

Note: Work in progress version! Preliminary, unfinished version in both content and language. Comments are welcome!

Abstract

In addition to fragmentation, polarization is one of the most established and discussed indicators of party systems. Although a lot of research has been done on this subject, studies vary concerning the specific operationalization. The common ground is that polarization conceptually represents some kind of aggregated ideological differences between parties. However, there is a significant disagreement about the specific measurement and, apart from that, most studies lack a substantial reasoning for their used indicator.

Based on the outlined concept of party system polarization, I examine common indicators: variance, standard deviation, mean absolute difference, range and counting extreme parties. I argue that each indicator has its specific pitfalls with respect to the concept. In addition, I outline possibilities of weighting and different approaches to determine party positions. Next, I compare these operationalizations on the empirical level. Referring to the outlined measurements of dispersion, the weighting functions and left-right approaches, there are 210 potential indicators of party system polarization. The analysis reveals that there are major disparities and varying measures lead to substantial conflicting results. Therefore, the choice of operationalization is decisive. Lastly, I discuss three, general suggestions based on the results.
Introduction

Ideological polarization is one of the most established and discussed indicators of party systems (Dalton 2008, Curini and Hino 2012, Sartori 1976). Various studies examine the effect of polarization on the output and outcome of political system (for a summary Curini and Hino 2012). For example, a high polarization presumable results in cabinet instability (e.g. Warwick 1992) and legislative gridlocks (e.g. Jones 2001). Furthermore, Dalton (2008) discusses polarization to be the reason behind the breakdown of several democracies, e.g. the Weimar Republic or the French Fourth Republic. However, despite its importance and the vast amount of studies in comparative politics, there is no consensus about operationalization. Studies rely on various different measurements (e.g. Gross and Sigelman 1984; Ezrow 2008; Mair 2001, Rehm and Reilly 2010, Best and Dow 2015) and a methodical examination of these indicators is almost completely lacking, even though the discrepancies seem to be critical. Different measures probably lead to substantial varying results.

Some studies indicate this. Curini and Hino (2012) examine the hypothetical effect of electoral systems and fragmentation on polarization and summarize some conflicting findings. For example, Dalton (2008) presents a relationship between the district magnitude and his variance-based measurement of polarization. In contrast, Budge and McDonald (2006) cannot identify any relationship between the ideological range and the electoral system, and so forth. Furthermore, Schmitt and Franzmann (2016) find differences concerning the relationship between the coalition type and three polarization indicators: the number of extreme parties, ideological range and vote-share weighted standard deviation. Here, center coalitions only have a centrifugal effect on range. In addition, Powell and Ascencio (2016) compare the “perceived” and “declared” polarization: In the former concept, party positions are measured by a voter survey, whereas in the second one party position are based on the Manifesto Project (MARPOR; Volkens et al. 2015). Apart from that, the authors compare polarization based on the RILE-Index (Budge 2013) and the Franzmann-Kaiser-Index (Franzmann and Kaiser 2006). The authors discover crucial differences between all three indicators (Powell and Ascencio 2016).

Referring to a cross-validation, Dejaeghere and Dassonneville (2015) find a high, but not perfect correlation (r = 0.8) between Dalton’s Index (Dalton 2008) and the unweighted standard deviation. Further, the Alvarz-Nagler-Index (Alvarez and Nagler 2004) correlates negatively with Dalton’s Index (r = -0.84). In addition, Rehm and Reilly (2010) show descriptively a substantial gap between polarization based on expert and voter survey data. For example, we can observe a significant and continuing centrifugal dynamic in Canada since 1980s referring to expert ratings. In contrast, the indicator rest upon voter surveys suggest a depolarizing trend since the 1990s. Finally, Best and Dow (2015) investigate disparities in polarization measures most detailed. To the best of my knowledge, this working paper is the only study focusing on the methodical examination and comparison of established polarization indicators. Here, the authors contrast measures based on dispersion and range. In accordance with the other studies, the authors observe substantial disparities between the standard deviation- and range-based approaches. They conclude that ideological range is most suitable to measure party system polarization (ibid.).

Hence, the outlined studies reveal crucial disagreements, but the present analyses are rather fragmented and still, there is a lack of comprehensive knowledge about the indicators. This problem is further compounded by two considerations: (1) there is a vast amount of measurements and polarization formulas. However, empirical studies have to make a reasonable choice between all these options. In the later analysis, I compare 210 different possibilities to measure party system polarization. (2) Besides the large variety of possible operationalization, most studies lack a substantial reasoning for their used measurement. Moreover, I argue that the general problem is the lack of a
coherent conceptualization of party system polarization. Because of this deficiency, a decision for a specific measure is hardly possible, so that most choices seem to be rather arbitrary. Hence, on the one hand, we assume that party system polarization is a decisive factor in understanding the functioning of democracies. On the other hand, comparative studies rely on various, conflicting measurements without knowing the particular extent and reasons for these discrepancies. Therefore, this paper examines the various polarization indicators by answering the following question: How can we measure valid party system polarization? Furthermore, is there a superior measurement? Therefore, I outline the utilized measurements and their variations. I argue that the vast amount of variations can be reduced to four elementary features: (1) the underlying measure of dispersion, (2) the usage of a weighting function, (3) the approach of measuring ideological positions and (4) the considered number of dimensions. Nevertheless, prior to a reasonable assessment of measure’s validity, a concept of party system polarization is needed. This fundamental step is tricky due to the common vague description of polarization in literature. In general, party system polarization is outlined as ideological differences (or dispersion) within the party system. In addition, discussed features are also the presence of extreme parties and the homogeneity of parties. However, I argue that party system polarization is basically an indicator of ideologically-based patterns of interaction between parties. Thus, patterns of contest and, as consequence of an increasing trend, of conflict occur in a polarized party system. On the contrary, cooperation between parties is largely absent. On this basis, I outline the (dis-)advantages of each measure and show empirical differences between them. Here, the discrepancies are remarkable high and, hence, the decision for a specific measure determines largely the result in quantitative analysis. Unfortunately, there is no superior measure. Each indicator captures partially the phenomena and fails under some circumstances. Nevertheless, the knowledge about the specific pitfalls of each measure is decisive for understanding the results of comparative studies. Furthermore, the most important consideration should be the way of measuring ideological positions. Here, the discrepancies between the indicators are especially large.

The Concept of Party System Polarization

Even though party system polarization is an established research subject, defining the concept is not an obligatory task, as one might surmise. In particular, the initial problem is that the wide range of studies pay largely little attention to conceptualizing party system polarization and a detailed conceptual definition is mostly lacking. Moreover, many studies actually only operationalize, but not conceptualize polarization (e.g. Lachat 2008). Further, several studies also use their operationalization to precise the concept. For example, Ladner (2004) defines polarization as the average distance to party system’s ideological center of gravity referring to the variance-based measurement of Taylor and Herman (1971). Based on such a concept, a comparison of the different measures would be superfluous because the variance-based operationalization is accurate by definition. Nevertheless, quantification should be the result and not the starting point of conceptualization (Sartori 1970: 1038), or in words of Gerring (1997):

“One must [...] have some idea of what one is looking for before one can find it. When concepts are defined "backwards" - by working out methods of measurement first - it may only complicate the task of social science inquiry since this encourages a rather facile approach to definition [...]“

Indeed, this non-conceptualizing (or measurement-based conceptual) approach would not be problematic if there is an undisputed and well-known defined concept of party system polarization. However, I argue that this is not the case. Especially, the concept of polarization contains some explicit
varying theoretical aspect and some undiscussed implicit axioms, which are not considered regarding the operationalization. Hence, one must state more precisely, what we are looking for. Nevertheless, we need an appropriate starting point to define the concept of polarization. Regarding the analysis of party systems, most studies refers to Sartori (1976) and his concept of polarization, e.g. Sigelman and Yough (1978), Knutsen (1998) or Pelizzo and Babones (2007). Originally, Sartori (1976) established polarization to distinguish multi-party systems and explaining dynamics of party competition. Despite the importance for his analysis, the conceptual introduction of polarization is initially rather short. Regarding a “more-than-one” party system, Sartori (1976: 126) defines polarization as ideological distance between parties:

“The term is used first to denote an ideological distance, that is, the overall spread of the ideological spectrum of any given polity [...]”.

In addition, Sartori (1976: 135) argues, “(we) have polarization when we have ideological distance (in contradistinction to ideological proximity)”. This view on polarization is adopted with little variation by most studies. Thus, Dalton (2008) defines polarization as “[...] the degree of ideological differentiation among political parties in a system.” Comparable definitions can be found, inter alia, by Powell and Ascencio (2016), Han (2015), Kim et al. (2010), Pardos-Prado and Dinas (2010) or Klingemann (2005). In these variants, the major disagreement is the use of the terms “ideological differences” or “ideological dispersion”. Up to this point, I will stick to the expression “differences” and dissolve this question later. Even though this basic definition is quite simple, the concept already contains the sub-concept ideology.

Thus, the conceptual core is based on somehow aggregated ideological differences within the party system. Nevertheless, we cannot elaborate the type of aggregation in more detail because of the current rather general concept. Up to this point, a high polarization implicates simply higher ideological differences between parties and vice versa. The nature of ideological differences, in which we are particularly interested, is still fuzzy.

This nucleus of polarization is often extended by the element of extremism (e.g. Pelizzo and Babones 2007: 56; Warwick 1992; King et al. 1990). Hence, a high polarization also implies the presence and electoral success of extreme parties. This consideration is already part of the original argumentation of Sartori (1976: 132) who introduced the concept of anti-system parties related to the type of polarized pluralism. Here, the dominant ideological dimension, which structures party competition, includes also a constitutional characteristic (ibid.: 335-340). According to this argument, parties positioned far off the ideological center evince an anti-constitutional attitude. In consequence, a higher polarized system should also contain a higher amount of anti-system parties. Recently, Capoccia (2002) has reassessed the concept of anti-system parties by detaching the concept from its historical reference to totalitarian parties in the 1960s and 1970s. Therefore, Capoccia (2002) defines relationally anti-systemness as the particular high ideological difference of a party to the other ones. Thus, anti-systemness implies the party’s ideological alienation from the established system.

This definition does not imply any assumption about the manner of alienation. Hence, such an anti-system party does not have to be authoritarian or anti-democratic (ibid.). Of course, this difference-based definition of anti-systemness fits well to our present core definition of polarization, because we do not have to integrate an additional sub-concept of “extremism” or “anti-system”. However, the

---

1 In short, my argument is that the concept of party system polarization contains ideological relations between parties, which characterize the occurring patterns of interaction. Therefore, the term “differences” is closer to the concept than “dispersion” or “distribution”.

4
view on ideological differences becomes more complex. Considering a relational alienation of parties in ideological polarization, we can expect a non-linear increase of polarization between a medium and a high difference. To be more precise, relational anti-systemness implies a theoretical threshold, when the alienation of a party begins. A change of parties’ differences that includes a passing of that threshold should be more important than a change, which does not imply any shift in parties’ alienation.

Sometimes, studies of polarization discuss a further, conceptual element: parties’ homogeneity. Most studies examining the ideological variance within parties seem to be present in the U.S. context, e.g. Poole and Rosenthal (2007) or Lee (2015). Especially, these studies use a specific indicator referring to the case of presidential two-party system. Here, polarization is commonly determined by the deviant voting behavior of parliamentarians. Thus, Han (2015: 3) argues that “conceptualisation of party polarisation can be different across party systems.” Han aims especially to distinguish between two- and multi-party systems. However, the argumentation seems to be problematic. On the one hand we can apply different operationalization because of the defined concept, but on the other hand we can apply different concepts depending on the context. In contrast, the definition of a concept depending on the context blurs the meaning of a used term.

As an exception in comparative studies, Rehm and Reilly (2010) apply also a polarization measure based on homogeneity referring to the concept of societal polarization (Esteban and Ray 1994). Therefore, they determine the homogeneity of a party due to its voters’ standard deviation on a left-right scale (Rehm and Reilly 2010: 45). In consequence, this measure combines common operationalization of party system and societal polarization. Nevertheless, I want to distinguish between both conceptual level – society and party system – and avoid a context-specific element in defining polarization. Thus, I concentrate on two basic elements of polarization:

1. Differences between parties or the distribution of parties
2. The presence of extreme or relative anti-system parties

Both elements depend on the sub-concept of ideology. Therefore, most studies of polarization consider a left-right dimension referring to spatial theory of party competition. Nevertheless, the conceptual meaning of left-right varies. On the one hand, the dimension can include a specific set of issues, which is invariant regarding the context, e.g. the economic left-right dimension of Downs (1957). On the other hand, the left-right continuum is also interpreted as “super issue” which summarizes all relevant issue positions within party competition (for a summary: Gabel and Huber 2000). Here, the underlying assumption is that valence as well as position issues can change over time and be different. Thus, the structure of party system is flexible and not given a priori. Additionally, there is a discussion about the number of relevant dimensions. Especially, one dimensionality is common criticized for being too simple to capture the complex structure of part systems. Nevertheless, studies of polarization mostly stick to a one-dimensional approach with only few expectations.

Originally, Sartori (1976) argues that the existence of a dominant ideological dimension is not necessarily given in a party system, but its existence is necessary for the possibility of polarization. If a dominant ideological dimension does not structure the party system, it will rather segment than

---

2 In these studies, polarization is usually measured by (DW-)NOMINATE-Scores. This indicator measures positions by the legislative roll-call voting behavior and a multi-dimensional scaling approach (Poole and Rosenthal 1985).
3 I summarize rather superficial the concept of ideology. Gerring (1997) outlines a very comprehensive overview.
4 For example, Andrews and Money (2009) measures range-based polarization in a two dimensional space. In addition, Spies and Franzmann (2011) study the effects of polarization on the electoral success of extreme right parties referring to two, different dimensions.
polarize (ibid. 216ff.). Here, the argumentation does not rest on a specific meaning of “left-right”. The idea is rather that any kind of ideological structure allows an ideological differentiation. Therefore, party system polarization presupposes some structured, dominant ideological dimension within party competition.

To the best of my knowledge, all studies of polarization treat the presence of dimensionality as axiom. In majority, one dimensionality is also a priori assumed. I did not find any test of dimensionality before measuring polarization in a comparative study. Nevertheless, I will further discuss dimensionality in the empirical section. Regarding the concept of polarization, the central argument is simply that some kind of structure is necessary for party system polarization.

Up to this point, I have outlined all relevant conceptual definitions and elements of party system polarization discussed in literature. Nevertheless, this issue seems to be still rather puzzling. To recourse to the original beginning of this section: What are we looking for regarding party system polarization? I argue that this question remains unanswered. To illustrate this point, a further, basic question may be useful: Why are we interested in ideological differences and alienation?

Referring back to Sartori (1976), the original purpose of the concept is the analysis of dynamics within party competition. Therefore, he criticize the previously common concentration on fragmentation, because, in contrast to polarization, this indicator does not reveal sufficiently accurate the structure of party competition (ibid. 199ff.; in addition Klingemann 2005: 38ff.ff.). Furthermore, a polarized system differs from a moderate system by the politics of outbidding, the irresponsible as well as anti-system opposition (Sartori 1976: 133ff.). In a polarized party system “[...] cleavages are likely to be very deep, (. ) consensus is surely low, and (. ) legitimacy of the political system is widely questioned” (Sartori 1976: 135). In addition, Vegetti (2014) argues that polarization implies rather political conflict than mere policy dispersion:

“If ideologies are defined as belief systems, then ideological polarization should imply a type of political conflict that spans across issue domains, where the sum is more important than the parts” (ibid.).

In other words, the conceptual idea is that polarization distinguishes party systems due to different ideological-based patterns of interaction.

The conceptual linkage between conflict and polarization is occasionally present in literature. For example, Steiner and Martin (2012) argue that party dispersion should be measured on the conflict dimension of party system. Further, Schmitt (2009) measures ideological conflict by a range-based indicator of polarization and Pardos-Prado and Dinas (2010: 767) reason that the dispersion-based measure of polarization “indicate the level of political conflict”, and so forth. Furthermore, especially studies of societal polarization highlight the connection between conflict and polarization. Thus, Rodrik (1999: 393) argues that the chance of finding a compromise between two groups is less in polarized contexts, or rather “it is difficult to coordinate on a “fair” distribution of resources”. In consequence, polarization causes conflicts, because groups perceive non-cooperative strategies as a more promising approach.

Nevertheless, the reducing of polarization on the contrast between conflict and cooperation is a bit shortcoming. On this view, we could only distinguish few party systems. In consequence, all cases with the absence of relevant conflicts would be equal. However, patterns of interaction between parties
are described more purposeful with four types of interaction: cooperation, negotiation, contest and conflict (Franzmann 2011: 319ff.; Bartolini 1999: 444). Bartolini (1999: 439ff.) outlines the differences between these types on several dimensions. Contest as well as cooperation presuppose a common goal, but the principal of contest is individualistic and other’s interests are not considered. In conflicts “actors enter into a social relationship in which they inflict damage on each other” (Bartolini 1999: 339). Here, a common goal is absent and, of course, the interests of others are undermined. In contrast, negotiation characterizes actors’ diverging goals, but the own interests are partially subordinated. Therefore, the principle is still solidarity which is the common feature of negotiation and cooperation (ibid.: 439-444) and, further, negotiation can be the starting point of cooperation. My argument is that the occurring patterns of interaction between two parties are partially determined by the ideological difference between them. First, an increasing, but moderate polarization is presumable characterized by patterns of contestation. Parties still accept the rules of the democratic game, but their ideological differences prevent the occurrence of cooperation. With increasing polarization, the minimal consensus erodes and, in consequence, conflicts between parties arise. Here, ideological differences cause a situation, in which any agreement is no longer possible, e.g. between democratic and authoritarian parties. Sartori (1976: 131ff.) describes this state by the politics of outbidding, an irresponsible opposition and the other consequences of polarized pluralism (in addition Schmitt 2014). In contrast, patterns of interaction are more complex referring to a lowering polarization. Ideological proximity is presumably not sufficient for the presence of cooperation (Franzmann 2011: 323). At least, goals referring to policy-, voter- and office-seeking affect additionally the possibility of cooperation. Even though two parties are ideologically similar, the competitive situation can lead to disagreeing calculi (e.g. Franzmann and Schmitt 2016). Nevertheless, following the outlined arguments, we can deduce a logical dependency between ideological differences and the occurring interaction: A certain degree of ideological proximity is necessary for cooperation as well as a certain degree of ideological distance is sufficient for conflict.

Figure 1: Ideological differences and patterns of interaction

Therefore, the interaction between two parties with a low or medium distance is characterized by varying patterns of interaction. However, the actors accept the same rules of competition. In the words

---

5 Originally, Bartolini (1999, 2000) understand cooperation as an antonym for competition. To avoid terminological inconsistency, I follow the proposal of Franzmann (2011) to distinguish between contest and competition. Therefore, Franzmann defines competition as “an institution in which parties strategically cooperate or contest as political actors to gain political power.” (ibid. 320). In contrast, contest is a type of interaction. Hence, competition is concept on macro level and contest on actor level.
of Sartori, we suppose a “competition on issue” and not a “competition on principle”. Within this “competition on issue”, cooperation as well as contestation are possible occurring interactions. With increasing differences, these patterns change. First, the probability of cooperation lowers, growing contestation occurs and finally, a fundamental conflict appears.

I argue that we are conceptually interested in these ideological-rooted patterns of interaction when we examine party system polarization. Especially the literature based on Sartori’s framework argues in that way. Therefore, I propose following conceptual definition of party system polarization:

**Party system polarization incorporates the aggregated ideological-based patterns of interaction within party competition – especially the determination of contestation and conflict by ideological differences.**

Based on this outlined, conceptual approach, I examine the proposed measurements of polarization in the next section. First, I compare their fundamental functionality and outline potential pitfalls. In the next step, I examine the empirical differences between these proposed operationalizations.

### The Measurements of Party System Polarization

Studies of polarization utilize many different formulas for aggregating polarization (e.g. Best and Dow 2015; Powell and Ascencio 2016; Rehm and Reilly 2010). Nevertheless, many proposals are quite similar. For example, Dalton (2008) suggests a measure of polarization based on standard deviation weighted by parties’ vote share. Actually, the only difference to other measures of that kind (e.g. Ezrow 2008) is the division by five. Because of this transformation being linear, Dalton’s indicator correlates perfectly. However, there are also essential disagreements within the proposed formulas. In the first step, I reduce these to the underlying measurement of statistical dispersion and the type of weighting. Therefore, I outline five formula types – variance, standard deviation (SD), mean absolute difference (MAD), range and counting the number of extreme parties – and I argue that each of them can further bases on three different weighting functions – equal, vote or seat share weighting.

The weighting function refers to the idea of parties’ varying relevance within the system. Based on this argument, ideological distances are included to a different extent (Ezrow and Xezonakis 2011: 1173; Alvarez and Nagler 2004: 50). Thus, patterns of interaction between less relevant parties are assumed less relevant with respect to the dynamics of party competition. This characteristic is usually taken in account by party’s vote or seat share. Here, the implicit assumption is that relevance is a linear function of one of the two shares. In consequence, a party with a vote share of 20 percent is twice as important as a ten-percent-party, and so forth. Under some circumstances, this assumption can be problematic. One might argue that party’s vote (or seat) share do not solely determine its relevance (Ezrow and Xezonakis 2011: 1173). Moreover, the specific institutional context and competitive situation shape presumably party’s relevance as well. Therefore, vote (or seat) share is a function of relevance but contains some error. Furthermore, it is reasonable that this error does not vary independent of the degree of weighting. There may be context specific thresholds, which imply a disproportional change of relevance. For example, when a party passes the five percent threshold in Germany, its relevance rises significantly. Many more examples are conceivable, e.g. referring to the relevance for coalition

---

6 However, it would also be pointless to analyze party competition dynamics with an indicator, which is not relevant for the types of interaction between parties.

7 Dalton (2008) propose following formula: \[ P = \sqrt{\frac{\sum_{i=1}^{N} v_i (p_i - p)^2}{N}} \]; \( v = \) vote share, \( p = \) party position and \( N = \) Number of parties.
Satori’s framework.

Evans (2002: 169) argues this approach to be the “wrong measurement” according to Satori’s framework. He outlines an example that contrasts a five-party system with a two-party system:

---

8 In addition, Dow (2001: 119) do not find any substantial difference between the aggregated, absolute and squared distances to ideological median of the party system.
The two-party system implies a vote-weighted SD-polarization of two (Eq.1b and weight 2). In contrast, the other one indicates only a value of 1.8. Nevertheless, the degree of ideologically conflicts is presumably higher in the right case. Here, two extreme parties are present, which are ideologically alienated from the established party system. Actually, this example is theoretically a classic starting point of polarized pluralism (Sartori 1976; in addition Schmitt 2014; Schmitt and Franzmann 2016). In contrast, the left case represents a common two-party system. Here, parties show a medium ideological difference and we can accept patterns of contestation between them. In this specific case, Evans’ critique of this operationalization seems to be valid because the presence of a substantial ideologically-based conflict should be indicated by the measurement.

This critique reveals a general characteristic of the SD-approach. Here, the degree of polarization depends highly on party system’s polarity. Because distances are measured in relation to the mean, party systems with actors positioned around the ideological center are characterized by a low or medium polarization. In consequence, a bipolar system shows usually higher polarization values compared with tripolar systems. This is also the reason why two-party systems shows empirically relative high patterns of polarization in cross-country comparison referring to this indicator.

Furthermore, this argument also highlights the circumstance, that the weighting of a party positioned at the center of gravity determines largely the polarization value. To be precise, a party – exactly positioned at the center – reduces the obtainable maximum value by \( w_i \times \text{max} \). For example, ideology is measured on an eleven-point left-right scale (0 to 10). Here, the maximum polarization indicated by equation 1b (and weight 2) is clearly five. If there is a center party with a vote share of 0.5, the still obtainable value is maximal 2.5 – regardless of the other parties’ vote share or position. Thus, a hypothetical system with three parties is medium polarized in following constellation: Party A is a maximal left-wing party and holds one quarter of the votes. Party B is the center party with a vote share of 0.5. Finally, Party C is a maximal right-wing party and holds the left vote share. In such a system, we could assume an irresolvable ideological conflict between all three actors and polarization should be very high to indicate this fact.

This example reveals also an implicit, undisputed assumption. Usually authors argue that maximal polarization is reached in a two-party system with a maximal extreme left- and right-wing party (e.g.  

---

9 The crucial point is whether there is an unequal or equal number of poles. If the number is unequal, there is supposable a group of parties clustered around the ideological center of gravity. This proximity to the center has a disproportionate negative influence on variance- and SD-based indicators.
Rehm and Reilly 2010: 43; Maoz and Somer-Topcu 2010: 812). This argumentation is incomplete under the consideration that an ongoing, ideological-based conflict between two parties arises before the maximal distance is achieved. To illustrate this argument, the following figure outlines two examples:

**Figure 3: The Maximum of Polarization – Two Answers**

Both cases represent highly polarized party systems, but the question is whether one (or both) example(s) represent(s) the theoretical maximum. The “classic” answer is illustrated by the right party system. Here, one continuous conflict is present. Nevertheless, I argue that an insuperable ideological conflict may arise between a center party and a maximal extreme right- or left-wing party. Thus, we would assume a “competition on principle” between all three parties in the left case (see figure 3). As consequence, there are three deeply ideologically-rooted conflicts and not any possible simple majority. Based on the outlined concept, both examples are, at least, equally polarized because of the characterizing ideological conflict – and the absence of other interaction patterns. Here, the SD-based measure fails, because it estimates a substantial, lower polarization in the left case.

Finally, there is another potential fallacy. The problem arises due to a skewed polarization trend. If the party system shows only a one-sided polarization dynamic, the indicators can lowers under some circumstances. This occurs when parties of the “skewed” side win also additional votes. To demonstrate this consideration, figure 4 outlines such a dynamic:

**Figure 4: The problem of skewness**

The starting point is the left party system. There are three moderate (A, B, C) and one extreme right-wing party (D). Next, party C as well as D wins additional votes. However, the weighted standard
deviation lowers because of the shifted center of gravity. This misleading trend of the indicator is possible even though the right-wing-party positions more to the right. An alternative, but similar formula operationalizes polarization by another measurement of statistical dispersion: the mean absolute difference (MAD).\(^\text{10}\) The utilization of this approach is considerably rarer in comparative studies. To the best of my knowledge, Gross and Sigelman (1984) were the first utilizing such a measure.\(^\text{11}\) Here, the implementation is problematic, because the weighted differences are summed and, as consequence, this indicator correlates highly with fragmentation. Other applications of this indicator do not weight the distances by vote or seat share (e.g. Klingemann 2005: 46). Nevertheless, the aim is to implement a weighting function that is independent from the number of cases. Here, my proposal is:

\[
P = \sum_{i=1}^{N} \left[ w_i \ast \sum_{j=1}^{N} \left( \frac{w_j}{1-w_i} \ast |p_i - p_j| \right) \right]
\]

The summed weights are always one – ignoring the cases \(i = j\). However, these cases can be omitted regarding the summed weights, because here the distance is always zero. Thus, the weighting function does not correlate with the number of parties. Referring to the conceptual definition, the advantage of this formula seem to be the more straightforward measure. The formula aggregates the patterns we are actually interested in – ideological differences between parties. Nevertheless, the correlation between SD and MAD is relatively high referring to random generated data\(^\text{12}\), but heteroscedasticity is present (appendix fig. 16). The generated data reveals a specific relationship between both indicators.

The reason for the gap between MAD und SD is the different treatment of party clusters. SD polarization lowers especially due to parties around the center of gravity. On the contrary, MAD lowers by any kind of present party clusters – even a cluster of extreme parties. Thus, MAD can decrease because of additional occurring extreme parties.

Figure 5: The problem of clustering

---

\(^\text{10}\) The mean absolute difference is normally calculated by following formula: \(MD = \frac{1}{n^2} \sum_{i=1}^{N} \sum_{j=1}^{N} |y_i - y_j|\)

\(^\text{11}\) Gross and Sigelman (1984) purpose following formula: \(P = \sum_{i} \sum_{j} \frac{w_i \ast |p_i - p_j|}{1-w_i}\). Further, the differences are weighted by seat share.

\(^\text{12}\) Regarding random generated data, I calculate correlations coefficients of approx. 0.84. The conditions of the procedure are the following: random number of parties (2 to 20) with a random vote share (but greater than 0) and a random position (0 to 10). Further 100,000 of such cases were randomly generated.
Figure 5 outlines this fact. In the right example, SD indicates a higher polarization in comparison to the right four-party system (left $P = 4$; right $P = 4.5$). In contrast, MAD claims a reverse order of polarization. Moreover, the polarization lowers from eight to approx. 6.33. Both party systems are presumably shaped by an ideological conflict and indicating a lowering trend is not in line with the theoretical concept. On the contrary, I argue that the ideological differences are rather intensified in the four-party system. The logical dependency between SD and MAD basis on this treatment of party clusters: An increasing value of the SD indicator is simply necessary, but not sufficient for an increasing MAD value. Therefore, the MAD inherits all described pitfalls of the SD indicator.

The presumable, second most used measurement is based on range (e.g. Mair 2001; Sørensen 2014; Best and Dow 2015). Therefore, this formula measure simply the maximal ideological distance between parties:

$$\text{Eq. 3a } P = \max(p) - \min(p); \text{ if } w_i = \frac{1}{N}$$

This approach does not include any vote or seat share weighting. To the best of my knowledge, there is no explicit implementation of the weighting function in the range-based indicator. However, there is a rather implicit method: Few studies measure the range between the two major parties (e.g. Sørensen 2014: 432, Powell and Ascencio 2016).

$$\text{Eq. 3b } P = |p_i - p_j| - \text{ the two parties (i and j) with the highest vote (or seat) shares.}$$

Because this approach is utilized in several studies, I include this variation in the later empirical analysis. Nevertheless, this weighting function obviously changes the entire logic of the indicator and I except only little correlation between eq. 3a and eq. 3b. Both types would only indicate the same trend, if minor and major parties show the same competition dynamics. However, theoretical approaches assume the opposite (Sartori 1976; Hazan 1995): A centrifugal trend of minor parties and a centripetal trend of major parties. Here, “vote share weighted” range fails to indicate the polarization trend. The more interesting indicator is the unweighted range. Even though this formula is quite simple, this approach avoids some problems. Regarding the outlined examples (figures 2, 3, 4 and 5), range predicts correct the polarization in contrast to SD and MAD. Nevertheless, the range shows other, crucial problems: (1) The range does not indicate any competition dynamic within the ideological minimum and maximum. Therefore, the extremisation of moderate parties is not regarded by range – even though interaction patterns shift presumably to conflict. The following figure shows an example:

Figure 6: The problem of dynamics within the range
In addition, the MAD does not indicate a polarization trend in this example, too. The total sum of differences is the same, if the range (and the number of parties) is constant. Thus, MAD and range fail to predict the increasing polarization dynamic in this specific cases.

The second pitfall of range considers the ignoring of vote- or seat shares. Even though extreme parties win substantial and, in consequence, patterns of dynamics shifts more and more to conflict, the range remain the same. Thus, this measure cannot capture major shifts in electoral competition.

Lastly, comparative studies rely on a further indicator. Especially studies of coalition bargaining (e.g. Warwick 1992; King et al. 1990) operationalize polarization by the number, vote (or seat) share of extreme parties. Therefore, this approach rely on a definition of “extremism”. Here, I define extremism based on party’s ideological position. Regarding an eleven point left-right scale (0 to 10), parties are defined to be extreme, when (1) their ideological position is equal or lower than two or (2) their position is equal or greater than eight. I choose this threshold based on the usual classification of this scale (e.g. Hazan 1995: 427). Nevertheless, such threshold are always rather arbitrary, but this specification does not hinder to analysis the general logic behind counting the number of extreme parties. To avoid any dependency on a certain scale, I define the extremism based on the relative distance to the center of the scale ($c$) referring to the theoretical maximum distance to center ($\max(d)$). Therefore, the indicator is defined in the following way:

$$\text{Eq. 4a} \quad P = \sum_{i=1}^{N} e_i \quad f(e) = \begin{cases} 1, & \text{if } |p_i - c| \geq \max(d) * 0.6 \\ 0, & \text{otherwise} \end{cases}$$

$$\text{Eq. 4b} \quad P = \sum_{i=1}^{N} e_i \quad f(e) = \begin{cases} w_i, & \text{if } |p_i - c| \geq \max(d) * 0.6 \\ 0, & \text{otherwise} \end{cases}$$

Equation 4a outlines the indicators when parties are equally weighted and, in contrast, the other measure (eq. 4b) takes vote or seat share into account. Obviously, the pitfall of this indicator is the missing consideration of any dynamic within the moderate space of party competition. Regarding the outlined examples, this indicator performs relatively well, when the aim is to distinguish party system with or without extreme parties. Thus, the more polarized system is correctly predicted in Evan’s example. Further, this measure also does not fail to indicating skewed polarization trends (figure 4). On the contrary, this indicator do not regard any differences between parties and, hence, is incorrect whenever extreme parties are absent. Competition dynamics, which do not affect the defined status of any party (non- vs. extreme), are not revealed. Therefore, this indicator especially fails to distinguish low and moderate polarized party systems.

In this section, I outline four types of measures and, on this basis, three types of weighting. Each indicator has its specific pitfalls and the summary (tab. 1) reveals the structure of problems. Of course, I cannot quantify measures’ potential failures, but the exemplary description shows, that the indicators have different weaknesses:

<table>
<thead>
<tr>
<th>Table 1: Potential pitfalls of aggregation strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bipolarity</td>
</tr>
<tr>
<td>Bipolarity</td>
</tr>
<tr>
<td>Variance</td>
</tr>
<tr>
<td>SD</td>
</tr>
<tr>
<td>MAD</td>
</tr>
<tr>
<td>Range</td>
</tr>
<tr>
<td>Extreme Parties</td>
</tr>
</tbody>
</table>
Up to this point, one final choice of measuring misses: the determination of party’s ideological position. There are a vast amount of studies discussing possibilities and methods of measuring party positions (e.g. Franzmann and Kaiser 2006; Gabel and Huber 2000; Rehm and Reilly 2010) and consensus is lacking. There are, inter alia, three common types: (1) voter surveys, e.g. the Comparative Study of Electoral Systems (CSES: www.cses.org), (2) expert surveys, e.g. the Chapel Hill Expert Survey (CHES; Bakker et al. 2014) and (3) studies of manifestos, e.g. MARPOR (Volkens et al. 2015).

In mass surveys, voters are asked to place national parties on a left-right scale and party position are aggregated by these replies. The disadvantages are, e.g., the possible response bias (Lo et al. 2014; Saiegh 2015). For example, voters tend to position “their” party near the center of scale, and so forth. Furthermore, the available data is relative rare, especially regarding time-series cross-section analysis (see for a detailed overview Rehm and Reilly 2010). In addition, this approach is usually limited to a one dimensional left-right space.

Alternative, studies use expert surveys to determine party positions. Mair (2001: 24) argues that the advantage is that “the judgements of experts - who are presumably intelligent, well-read and informed - they acquire a certain weight and legitimacy.” Thus, these positions may be less biased than voters’ left-right placements. In addition, different studies show high correlations between expert and parliamentarians placements (for a review Franzmann 2015: 148f.). Moreover, it may be argued that expert responses combines “what parties say and what parties do” (Netjes and Binnema 2007: 42; in addition Mair 2001: 21) in one indicator.

Regarding comparative analysis the disadvantage is rather the limited available expert data (Rehm and Reilly 2010). A benefit is the possible application of a multidimensional approach: For example, the Chapel-Hill dataset include an item referring to general and economic left-right as well as the GAL-TAN dimension (Bakker et al. 2014: 144).

Lastly, researchers utilize party manifestos to deduce positions. Analyses based on the MARPOR are most common. Here, authors propose different approaches to determine ideological positions:

- RILE-Index (Laver and Budge 1992): The RILE-Index is the original left-right indicator of the MARPOR and presumably the most used one (Mölder 2015: 39). Here, coding categories are defined a priori to be left or right and combined to the scale (Budge 2013; for a review Mölder 2015). Further, the index does not integrate about half of the categories and the meaning of left-right is assumed to be invariant. In addition, a common point of critique is its rather low validity (Mölder 2015; Gabel and Huber 2000).
- Franzmann-Kaiser-Index (abbr. FK; Franzmann and Kaiser 2006): In contrast to the RILE-Index, Franzmann and Kaiser (2006) propose a step-by-step approach. For each election, all categories are classified as valence or position issues. Then, the relevant parties are selected and the context-specific meaning of the category is defined.
- Jahn-Index: Based on Norberto Bobbio’s theory of Left and Right Jahn (2011) identifies a core set of left-right categories. Further, he calculates the weight and meaning of the category within his core index based on multidimensional scaling. In addition, Jahn also calculates a “plus index” that basis on the remaining categories. These statements can vary regarding its specific meaning (ibid.).
- Further approaches are: (1) Elff (2013) determines party positions within an multidimensional, policy space based on a latent, dynamic state-space model. (2) König et al. (2013) rely on a Bayesian model. They define a priori two, contrary poles and assign the MARPOR coding categories to one of the poles. Further, they add additional information to the model based
on the CHES expert survey and the Euromanifestos Project. (3) Lastly, Lowe et al. (2011) calculates logarithmic scales on several dimension.

In the empirical analysis, I use following approaches and data sources to determine parties’ ideological position:

Table 2: Approaches to determine party positions

<table>
<thead>
<tr>
<th>Manifesto studies</th>
<th>Expert surveys</th>
<th>Voter Surveys</th>
</tr>
</thead>
<tbody>
<tr>
<td>MARPOR</td>
<td>CHES</td>
<td>CSES</td>
</tr>
<tr>
<td>– RILE (one dimensional)</td>
<td>– one and two dimensional</td>
<td>– one dimensional</td>
</tr>
<tr>
<td>– FK (one and two dim.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>– Jahn (one and two dim.)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>– König (one dim.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>– Elff (one and three dim.)**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>– Lowe (one and 21 dim.)**</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The two dimensional Jahn-Index is defined by the core- and plus-Index.
** Elff and König et al. present originally multi-dimensional approaches of measuring party positions. I define the one-dimensional index by calculating the mean value.

So far, all outlined formulas (eq. 1 to 4) are restricted to a one-dimensional approach. Nevertheless, I want to include the multi-dimensional positional data and test the effect of dimensionality: For example, are there relevant differences between one- and two-dimensional polarization? Therefore, the distance between two points is measured as Euclidean distance in an n-dimensional space:

\[ D(x_i, x_j) = \sqrt{\sum_{d=1}^{N} (x_{di} - x_{dj})^2} \]

Furthermore, I adjust the previous, one-dimensional formulas in the following way:

**Eq. 6a** (in contrast to Eq. 1a) \[ P = \sum_{i=1}^{N} [w_i \ast D(p_i, \bar{p})] \quad \bar{p} = \sum (w_i \ast p_i) \]

**Eq. 6b** (in contrast to Eq. 1b) \[ P = \sqrt{\sum_{i=1}^{N} [w_i \ast D(p_i, \bar{p})]} \quad \bar{p} = \sum (w_i \ast p_i) \]

**Eq. 7** (in contrast to Eq. 2) \[ P = \sum_{i=1}^{N} [w_i \ast \sum_{j=1}^{N} \left( \frac{w_j}{1 - w_i} \ast D(p_i, p_j) \right)] \]

**Eq. 8a** (in contrast to Eq. 3a) \[ P = \max(D_{ij}) \]

**Eq. 8b** (in contrast to Eq. 3a) \[ P = D_{ij} \]

**Eq. 9** (in contrast to Eq. 4) \[ P = \sum_{i=1}^{N} e_i \quad f(e) = \begin{cases} 1, & \text{if } d(p_i, c) \geq 0.6 \ast \max(D) \\ 0, & \text{otherwise} \end{cases} \]

In this section, I outline several elements of the operationalization of polarization – aggregation formula, weighting function and measure of ideological position. Therefore, measuring polarization requires a couple of methodical decisions:
Based on the outline data sources and aggregation formulas I calculate the different polarization indicators on empirical level. When CSES, MARPOR and CHES provides positional data, 210 different ways of calculating polarization are included in analysis: \((9 + 5) \times 3 \times 5 = 210\). However, in many cases, only some of the indicators are available.

Because the main purpose of the empirical analyses is to examine differences between the measure of polarization and not testing hypothesis, I include as many countries and elections as possible – with one exception. Countries are only included, if more than one approach of measuring party positions is available. In some countries, there is only positional data based on RILE (e.g. south Africa) or CSES (e.g. Hong Kong).

Furthermore, I calculate z-transformed values of polarization variables because the positional data varies regarding its scale and I am not interested in the variance based on scale differences. In the next section, I outline the empirical analysis.

---

13 Referring to this cases selection, the sample includes Australia, Austria, Belgium, Bulgaria, Canada, Czech Republic, Denmark, Estonia, Finland, France, Germany, Great Britain, Hungary, Iceland, Ireland, Israel, Italy, Japan, Latvia, Lithuania, Luxembourg, Netherlands, New Zealand, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey and the United States.

14 Because the positional data of some MARPOR approaches (Elff, König et al., Lowe and Jahn) has no defined maximum and minimum value, I determine the maximal distance to the center (see eq. 4 and 9) based on the statistical distribution: \(\max(D) = \max |p_i - \bar{p}|\).
The Empirical Analyses of the Measurements

Up to this point, I outline several issues related to the established indicators of polarization. Each measure has its specific pitfalls and, further, they may overlap with each other, but I also expect biased differences between them. However, I have only discussed them only on a conceptual level so far. Thus, a pragmatic question remains: Are there also relevant empirical disparities? To be straightforward, the answer is yes. Nevertheless, the interesting point is not the finding of differences. On the contrary, the astonishing result is extent of the gap between these indicators. Furthermore, the deviation is not only biased, the bias is context-sensitive. Therefore, we observe complex patterns of differences and, in consequence, the choice of operationalization is decisive regarding the findings in comparative analyses. Or, in another words: the confirmation (or falsification) of common hypotheses just depends on choosing the “right” indicator. To outline this answer in more detail, I analyze the differences step-by-step and reveal the important choices in operationalizing polarization. I start with the simple observation of differences. Therefore, I correlate all 210 indicators with each other and we can observe a (nearly) normal distribution.

Figure 8: Correlation between different polarization indicators

Figure 8 illustrates that there is a wide range of correlation coefficients. Thus, some indicators are very similar, but the majority of correlations is rather weak. The median is below 0.3. This result is already disillusioning. It is hardly conceivable, that all these indicators measure the same latent construct. In addition, some measures even correlates negatively with each other. This is rather the exception, but we can still observe 2619 negative correlations (12.6 percent). Hence, there are striking differences between the indicators. Nevertheless, the point of interest is the reason for this gap. Starting with illustrating descriptive, some already discussed suggestions of differences and similarities are empirically underpinned.

In the first step, I examine differences between the underlying measures of dispersion. Here, I outline five possible choices in operationalization: (1) variance, (2) standard deviation, (3) mean absolute difference, (4) range or (5) counting of extreme parties. The following figures illustrate the covariance between them:
First of all, the variance- and SD-based polarization are, as expected, quite similar and can be described by a function of a degree two polynomial with a very small error. The choice between variance and standard deviation is rather irrelevant. All other patterns of covariance are more interesting.

The most non-linear covariance can be observed on the bottom right – the covariance between the presence of extreme parties and SD. The logic of counting extreme parties differs significantly from the other measures of dispersion. The plot shows that the parallel occurrence of a low or medium SD in combination with a relative high number of extreme parties occurs in several cases. This circumstance is present, when on the one hand the fragmentation is high and on the other hand a cluster of moderate parties is positioned around the center. To give an example, these pattern can be observed in the Israeli election of 1984.

---

15 This errors solely basis on the standardization. The unstandardized indicator presents a perfect binary quadratic relationship.
Figure 10: Israeli Election of 1984*

*Parties from left to right: Democratic Front for Peace and Equality (HaHazit HaDemokratit LeShalom VeLeShivion), Progressive List for Peace (HaReshima HaMitkademet LeShalom), Movement for Civil Rights and Peace (Hatnuah Lezhiot Ha'ezrach), Change (Shinui), The Consolidation (Likud), Alignment (HaMa'arakh), Movement for the Heritage of Israel (Tnu'at Masoret Yisrael), Together (Yahad), Religious Torah Front, Sfarad's guards of the Torah (Shomrei Sfarad), Heritage (Morasha), Tzomet Crossroads (non-aligned movement for Zionist Renewal), Tehiya-Bnai (Tehiya-Bnai), Courage (Ometz) and National Religious Party (Miflaga Datit Leumit).

Figure 10 indicates the ideological party position (blue circles) based on the Franzmann-Kaiser Indicator. Because of the high vote share of Likud as well as HaMa'arakh, the vote-weighted SD is moderate (1.6). Nevertheless, the party system is remarkably fragmented and a high amount of the small parties is positioned far off the ideological center. Here, the SD indicator conceals the patterns of polarization. In contrast, the ideological range indicates a nearly maximal polarized system (7.5), which, on the contrary, may overestimates the present polarization in this case.

The covariances between the indicators based on SD, MAD and range are to some extent linear. Nevertheless, the scatterplots also reveal considerable discrepancies. Referring to the covariance between MAD and SD, most differences can be explained by the outlined divergent treatment of party clusters. Therefore, the MAD is especially effected by sample’s treatment of party alliances. Including the alliance as one actor or the single parties crucially changes the MAD value. In consequence, the usage of a MAD measure requires a reasonable and consistent treatment of such alliances.

In addition, the empirical data shows a further reason for deviating trends in MAD and SD. High differences occur in disproportional two-party systems. For example, two parties are included in the MARPOR sample regarding the Turkish election of 1951: the Democratic Party and the Republican People’s Party. The first wins approx. 60 percent of the seats. In consequence, the center of gravity is closer to this party and its distance to the mean is lower. On the contrary, range and MAD does not respond to such a disproportionality. Thus, the polarization may be better captured by the range or MAD in this specific competitive situation. However, these cases are very rare.

Furthermore, the patterns between range- and SD-based indicators are conform to the discussed discrepancies. Especially, the fact that the range does not capture competition dynamics beyond minimum and maximum, leads to the gap. Further, the highest amount on discrepancy is based on vote (or seat) share-weighted range indicators. Referring to the Israeli example, the range between the two parties with the highest vote share is extremely low \((P = 0.2)\). Thus, this variable indicates a very unpolarized system and is not valid in this case.
Up to this point, the empirical analysis already reveals some disagreements between the indicators of polarization. Nevertheless, these discrepancies may be crucial, but there are further varying elements: (1) weighting type, (2) dimensionality and (3) left-right measures. In the next step, the differences between the weighting types are outlined. These covariances reveal less interesting patterns.

Figure 11: Covariance between weighting types*

Comparing seat and vote share weighted indicators, the agreements between both versions is relatively high (top half, figure 11). Differences occur primary due to three facts: (1) In party systems with a high disproportionality the seat share indicator is usually lower because minor parties are excluded. Thus, examples for deviation are especially majority voting systems. (2) In party systems with electoral alliance, MARPOR assigns the vote share to the party and the seat share to the electoral alliance, or, sometimes, vice versa. The factor assigned to the electoral alliance reveals usually the lower polarization value. (3) The last reason of disagreement is a bit more tricky: The vote (or seat) share weighted range indicator measures the distance between the two major parties, but the determination of the second major party can change between seat and vote share. In consequence, the polarization levels are different between both weighting types.

However, the choice between vote or seat share is rather irrelevant in the majority of cases ($r = .97$). In contrast, the correlation between equally and vote share weighted polarization is remarkable lower ($r = .76$). Of course, the difference increases simply due to the disproportional dispersion of votes. Especially, we can observe differences in fragmented party systems like in the example of the Israeli election of 1984 (see fig. 10). Nevertheless, the scatterplot does not reveal another reason: The differences can be observed especially between the range-based indicators. The unweighted and vote share “weighted” range correlates only on a medium level ($r = .47$).
Next, I examine differences in measuring polarization due to the number of dimensions. Only, some left-right measures allow such a comparisons. Exemplary, differences between the FK- and CHES-indicator are outlined. Thus, the figure presents a comparison between one and two dimensionality:

**Figure 12: Covariance between one and two dimensional approaches***

![Covariance between one and two dimensional approaches](image)

* Z-transformed polarization values are included. Shortcuts: RNG = range, V = variance, WU = equally weighted, WV = weighted by vote share.

Both scatteplots on the right half of the figure show differences referring to the MARPOR based on the FK indicator. Here, one and two dimisional polarization correlate relatively high (above 0.9). Thus, in most cases there is no disagreement in evaluating polarization due to the included number of dimension. Nevertheless, there are some outliers. In contrast, the differences are remarkably higher referring to CHES data. Here, the correlation is only 0.7 and 0.59. Of course, the limited number of cases constraints the interpretation but the discrepency between expert and manifesto data is notable. This gap may be present due to the vary in computation of dimensions. One or more dimensional MARPOR indicators only vary with regard to the splitting of the same information source. On the contrary, CHES includes three different questions – overall, economic and GALTAN left-right party position. Thus, the overall left-right placement does not necessarily contain the same information like the other items – this may depend on the expert decision.

Lastly, I examine differences between varying meaures of ideological party positions. Here, the discrepancies are the largest.
Exemplary, the figures outline the covariance between polarization values based on some, varying left-right measures. Nevertheless, these patterns can also be observed between the other left-right indicators. The plots show some weak positive correlation. However, the plots reveal that the choice of the left-right indicator effects the polarization value decisive. Interestingly, the correlation between the left-right positions on party level is higher than between the left-right polarization. For example, Pearson’s coefficient is .66 between RILE and FK on party level. In contrast, the correlation between vote share weighted SD based on FK and RILE is only .49. Hence, the process of aggregation rather enhance discrepancies.

Up to this point, I have analyzed the covariance between different elements of polarization rather exemplary. Utilizing an ANOVA, I try to identify the crucial elements in operationalizing polarization in more detail. I do not use the analysis to evaluate hypotheses. Here, I am interested in the decomposition of the polarization variance. Of course, one could theoretically decompose the entire variance in its parts. However, there are technical constraints regarding the possible matrix size. Thus, I only include two-level interactions in varying models. However, I test several combinations and the outlined models represent the most interesting results.

To decompose the variance, the dataset is clustered on three levels (country, election, polarization measure) and includes four variables (applied measure of left-right, included number of dimensions, weighting function and measurement of dispersion). Due to this structure, the number of cases increases remarkably. The ANOVA models reveal following results:
First, the remarkable finding is the low partial explained variance of the context variables – country and election date. This ratio also does not vary substantially in different models. The partial R² suggests that only about 18 percent of variance is independent from the choice of measurement. Of course, this result illustrates the relevance of operationalization. To corroborate this finding, I calculate also a multi-level analysis with two level (country and election). Here, the intra-class correlation of 0.29 (election level) and 0.12 (country level) is revealed.

Further, the models show variance patterns in accordance to the outlined covariance plots. Most variance between the polarization indicators can be explained by the left-right measurement – especially in interaction with the specific election. Thus, choosing a left-right indicators changes the indicated polarization context sensitive. This is the most decisive choice in operationalizing polarization. Further, the context bias suggests that choosing a non-valid indicator results in non-interpretable and unknown biased polarization values.

Thereafter, the measurement of dispersion becomes relevant. This element of operationalization also interacts with the date variable and can partially explain six percent of the variance. In addition to the outlined models, there is also a three-level interaction effect between measurement of dispersion, left-right measurement and date. Furthermore, the weighting function and dimensionality reveal a substantial lower partial explained variance. As already outlined, these elements are only in few cases relevant and mostly the differences are not crucial.

These patterns are also supported by a cross-correlation of all 210 indicators. Here, the results are exemplary outlined by the following figure. The smaller cut-off represents perfectly the overall picture:

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sq.S</td>
<td>P.R²</td>
<td>Sig.</td>
<td>Sq.S</td>
<td>P.R²</td>
</tr>
<tr>
<td>Country (C)</td>
<td>938.8</td>
<td>.013 ***</td>
<td>696.7</td>
<td>.009 ***</td>
</tr>
<tr>
<td>Election date (E)</td>
<td>12225.7</td>
<td>.165 ***</td>
<td>11233.9</td>
<td>.151 ***</td>
</tr>
<tr>
<td>Dimensionality (D)</td>
<td>0.0</td>
<td>.000</td>
<td>0.2</td>
<td>.000</td>
</tr>
<tr>
<td>Left-Right Type (L)</td>
<td>363.5</td>
<td>.005 ***</td>
<td>190.1</td>
<td>.003 ***</td>
</tr>
<tr>
<td>Dispersion Measure (M)</td>
<td>27.3</td>
<td>.000 ***</td>
<td>27.3</td>
<td>.000 ***</td>
</tr>
<tr>
<td>Weight Type (W)</td>
<td>1.0</td>
<td>.000</td>
<td>0.7</td>
<td>.000</td>
</tr>
<tr>
<td>D * C</td>
<td>30.2</td>
<td>.000 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D * E</td>
<td>1211.6</td>
<td>.016 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L * C</td>
<td></td>
<td></td>
<td>608.0</td>
<td>.008 ***</td>
</tr>
<tr>
<td>L * E</td>
<td></td>
<td></td>
<td>23531.5</td>
<td>.317 ***</td>
</tr>
<tr>
<td>M * C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M * E</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W * C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W * E</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Model | 23371.7 | .315 ***| 47786.8 | .644 ***| 26818.1 | .362 ***| 24839.1 | .335 ***|
Residual | 50877.8 | 26462.6 | 47431.4 | 49410.4 |
Total | 74249.4 | 74249.4 | 74249.4 | 74249.4 |
N | 74046 | 74046 | 74046 | 74046 |

* Sig. Level: 0,001 ***; abbreviation: Sq. S. = Sum of Squares, P.R² = partial explained variance.
The figure shows that correlations between measures of same left-right indicator are relatively high. Most conspicuous is the deviation of the extreme-based indicators and weighted range. Here, the coefficients are relatively low. On the contrary, the other measures are quite similar – on the condition that they bases on the same left-right indicator. In accordance to the previous results, the correlation between different left-right types are only medium. Furthermore, the lowest correlation can be observed between extreme-based polarizations with any other indicator based on a different left-right measure.

Finally, I want to outline the relevance of these disparities for testing hypotheses. In terms of explaining polarization, two standard hypotheses are the effect of fragmentation and electoral system on polarization (e.g. Curini and Hino 2012). Here, fragmentation is measured by the common indicator of the effective number of parties (Laakso and Taagepera 1979) and electoral system are distinguished by majority, plurality and mixed systems based on the classification of Golder (2005). The reported coefficients of the electoral systems are the effects of the majority system referring to plurality systems. I calculated a pooled, linear regression analyses based on the z-transformed polarization values.
How to Measure Ideological Polarization in Party Systems (Schmitt, ECPR Graduate 2016)

Figure 15: Results of linear regression (DV: Polarization)

Table 4: Explaining Variance of Regressions Results: Regression Coefficients and Explained Variance*

<table>
<thead>
<tr>
<th></th>
<th>DV: R² (ENP)</th>
<th>DV: b (ENP)</th>
<th>DV: R² (Electoral System)</th>
<th>DV: b (Majoritarian System)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sq. Sums</td>
<td>Partial R²</td>
<td>Sig.</td>
<td>Sq. Sums</td>
</tr>
<tr>
<td>Dimensionality: D</td>
<td>0.000</td>
<td>0.002</td>
<td></td>
<td>0.000</td>
</tr>
<tr>
<td>Left-Right Type: L</td>
<td>0.025</td>
<td>0.095</td>
<td>***</td>
<td>0.881</td>
</tr>
<tr>
<td>Dispersion Measure: M</td>
<td>0.015</td>
<td>0.057</td>
<td>***</td>
<td>0.419</td>
</tr>
<tr>
<td>Weight Type: W</td>
<td>0.001</td>
<td>0.005</td>
<td></td>
<td>0.014</td>
</tr>
<tr>
<td>D * L</td>
<td>0.001</td>
<td>0.005</td>
<td></td>
<td>0.003</td>
</tr>
<tr>
<td>D * M</td>
<td>0.001</td>
<td>0.003</td>
<td></td>
<td>0.004</td>
</tr>
<tr>
<td>D * W</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
<td>0.001</td>
</tr>
<tr>
<td>L * M</td>
<td>0.087</td>
<td>0.324</td>
<td>***</td>
<td>0.396</td>
</tr>
<tr>
<td>L * W</td>
<td>0.008</td>
<td>0.031</td>
<td></td>
<td>0.086</td>
</tr>
<tr>
<td>W * M</td>
<td>0.062</td>
<td>0.231</td>
<td>***</td>
<td>0.320</td>
</tr>
<tr>
<td>Model (Between)</td>
<td>0.221</td>
<td>0.827</td>
<td>***</td>
<td>2.452</td>
</tr>
<tr>
<td>Residual (Within)</td>
<td>0.046</td>
<td>0.086</td>
<td></td>
<td>0.049</td>
</tr>
<tr>
<td>Total Squared Sums</td>
<td>0.267</td>
<td>2.538</td>
<td></td>
<td>0.413</td>
</tr>
<tr>
<td>N</td>
<td>207</td>
<td>207</td>
<td></td>
<td>207</td>
</tr>
</tbody>
</table>

* Sig. Level: .05 *, .01 **, .001 ***. Shortcuts: Sq. S. = Sum of Squares, P. R² = partial explained variance.
The scatterplots show that each potential hypotheses can be confirmed by some measure. Thus, the choice of operationalization effects remarkably the results of analysis. Nevertheless, strong effects of the electoral system and fragmentation on polarization are rather outliers. Further, fragmentation reveals more positive than negative effects and, on the contrary, majority systems rather lowers polarization. In addition, these effects do not vary unsystematically. For example, regarding the coefficients of ENP, all negative outliers are based on MAD. Here, the clustering of highly fragmented systems leads to the negative effects.

Furthermore, the variance is decomposed more systematically in four ANOVA models. All model confirms previous results: The decisive element is the choice of left-right. In addition, there are also remarkable interactions between measurement of dispersion and the left-right indicator as well as the weighting function. The first interaction (“M*W”) especially depends again on the weighted range. Here, difference have already been outlined and this varying logic leads to different results. Finally, the dimensionality is the most uninteresting element referring to its effect on the results. Here, the ANOVA does not reveal any relevant extent of between-variance. Finally, I discuss these results in the next section.

Discussion

A vast amount of comparative studies examines effects of (or on) party system polarization and the number of potential indicators is almost equally great. These studies present a wide range of very interesting findings in various fields of comparative research and polarization is one of the main indicators of party systems. Nevertheless, the literature also shows several conflicting results and comprehensive methodic analyses of the applied measurements misses largely. Hence, the impact of choosing a specific indicator is uncertain. In addition, the evaluation of measure’s validity is made difficult because of the mostly superficial defined concepts.

I argued that party system polarization represents the ideological-based patterns of interaction within a party system. Each proposed measurement – variance, standard deviation, mean absolute difference, range or counting extreme parties – has its specific strengths and weaknesses regarding to this concept. In addition, the empirical analyses shows enormous differences between the 210 operationalized indicators. Hence, the choice of a specific indicator is not only relevant on a conceptual level, but it also effects heavily the findings of comparative studies. All kinds of potential hypotheses could be “verified” by some indicator of polarization. This finding may be disillusioning, but the outlined analyses also suggest potential solutions.

Referring back to the original research question: How can we measure party system polarization? Based on the proposed concept, I suggest the following strategy:

1. Comparative studies should use, at least, three, different indicators: standard deviation, range and counting extreme parties. The advantage of this combination is that these measures differ regarding their specific failures and the phenomena of party system polarization is widely captured by this combination. Of course, this result is only preliminary. It would be preferable to develop an indicator that combines the measures’ advantages and avoids the outlined disadvantages. Up to this point, there is no superior measurement of polarization.

2. The sample should only include relevant parties based on a reasonable concept. One might surmise that this argument is trivial, but most studies do not pay any attention to this topic and apply the sample of the original data source. Each included party, which is not relevant with respect to the patterns of interaction within party competition, bias the measure of polarization. Thus, an additional effort should enhance significantly the validity of the results.
Furthermore, a reasonable sample lowers the necessity of weighting parties by vote or seat share, which has interpretation problems.

3. The last, but most important point is that each study of polarization should contain a justified concept of ideology and a substantial reasoning for choosing a specific type of measuring party positions. Here, the variance between the indicators is too large to ignore this problem. In addition, non context-sensitive measures may be a priori non-valid indicators referring to the concept of polarization. When we are interested in any kind of possible ideological conflict, the measure should be flexible in including different issues depending on the specific structure of the analyzed party system. Further, it seems plausible to test the presence and extent of dimensionality before measuring polarization. However, this seem to be only relevant in a few cases referring to the empirical analyses.

Appendix

Figure 16: Differences between MAD- und SD-Polarization

References

How to Measure Ideological Polarization in Party Systems (Schmitt, ECPR Graduate 2016)


