Navigating Polarization: Insights from Behavioral Science

Kati Kish Bar-On¹  Eugen Dimant  Yphtach Lelkes  David G. Rand
MIT  University of Pennsylvania  University of Pennsylvania  MIT

Abstract
Polarization is a critical issue in today's world, impacting democracy and sparking debates on climate change, populism, and global health. In this perspective, we explore polarization's nature, origins, and measurement challenges. We discuss how different disciplines and perspectives define polarization, either as an individual mindset or a collective state. Measuring polarization is complex, with various methods offering different insights. Our discussion covers interdisciplinary views, measurement approaches, and the limitations of existing research. We conclude by looking at future directions for understanding and addressing polarization.

1. Introduction
For centuries, scholars and researchers have explored the dynamics of polarization. Historical analyses reveal that the phenomenon has characterized the US Congress for much of its history (Brady and Han, 2007). Outside the US, Karl Marx and Max Weber were two of the earliest and most prominent figures to discuss class polarization. Class polarization refers to the division of society based on different social classes, often associated with disparities in income and wealth. Party polarization, on the other hand, occurs when political parties move towards extreme ideological positions, creating a significant divide within the political landscape. Throughout the 1970s and 1980s, the concept of party polarization underwent a process of evolution and expansion, extending its influence across diverse fields of study and spawning various iterations. In psychology, the concept of group polarization emerged, while in economics, the notion of social polarization gained popularity (Pahl, 1984; Moscovici and Zavalloni, 1969). By the early 2000s, research on affective polarization began to flourish, emphasizing the role of groups and identities as fundamental elements in American politics (Iyengar et al, 2012).

¹ Inquiries can be sent to: katik@mit.edu

Electronic copy available at: https://ssrn.com/abstract=4784194
Affective polarization research has its roots in the United States, with rich documentation of its presence and impact (Iyengar et al., 2019). Recently, the scope of this research has widened to include studies on partisan antagonism in various Western democracies (Boxell et al., 2022; Röllicke, 2023; Turner-Zwinkels et al., 2023). The emergence of far-right political parties in numerous European nations has evoked intense aversion among established political entities (Gidron et al., 2019; Helbling and Jungkunz, 2020; Mudde and Kaltwasser, 2018). Countries such as France, Italy, and Germany have experienced political upheavals due to these radical parties, while Spain, the United Kingdom, and Canada grapple with separatist movements within their political frameworks. Scholars suggest that American partisan hostility ought to be interpreted in a broader, comparative context (Adams et al., 2020). Although the initial studies of affective polarization in Europe have enriched the understanding of the phenomenon, there is an ongoing debate regarding the need for more refined measurement tools, advanced theoretical frameworks, and a clearer depiction of affective polarization's significance (Wagner, 2024). Our forthcoming analysis seeks to address these debates and investigate them through a multidisciplinary lens.

2. The complexity of Polarization

2.1 Interdisciplinary Perspectives: Definitions and Dynamics

Polarization is defined differently across various disciplines, and each brings its unique perspective to understand its causes, manifestations, and consequences. A plethora of terms such as identity, group dynamics, extremity, tensions, trust, values, biases, beliefs, norms, emotions, misperceptions, epistemic conflicts, partisanship, stereotyping, cohesion, diversity, and prejudice are employed to articulate the nuances of polarization. Though numerous theories address these aspects individually (Hale et al., 2020), we do not yet have a singular, all-encompassing theory of polarization. This may stem from the absence of a consistent ontological framework to construct such a theory. What are the foundational elements that constitute polarization? What is the set of entities that we can put into boxes to draw arrows between; that we can imagine interacting with other entities that cause, reduce, or affect polarization?

The nature of polarization is multifaceted, encompassing a wide range of constructs and metrics. Some theories posit that polarization results from individuals' maintaining their social identity within groups (Abrams et al., 1990), while others view it as the degree of social tension predicated on individuals' feelings of inclusion or exclusion within group dynamics (Permanyer, 2012). Economists traditionally define polarization as a societal characteristic marked by distinct groups that are internally cohesive yet distinctly different from each other (Esteban and Ray, 1994; Wolfson, 1994). Political scientists interpret polarization as the shift in political attitudes toward ideological extremes, and cognitive scientists examine how cognitive biases and information processing contribute to the entrenchment of extreme
beliefs. Thus, the term "polarization" refers to a spectrum of distinct social configurations and dynamics, each with specific traits.

However, the complexity of the term extends beyond its varied interdisciplinary interpretations. Within any single field, there is often ambiguity surrounding the exact social phenomena related to belief formation being referred to or the manner in which opinions are polarized. Ideological polarization often denotes divergence in political parties' stances on societal, economic, and political matters, whereas affective polarization generally indicates an emotional aversion and distrust toward political adversaries. Although theoretical and empirical distinctions between these concepts have been proposed (Lelkes, 2018; Iyengar et al., 2019), some studies show a correlation between ideological and affective polarization (Abramowitz and Webster, 2016; Rogowski and Sutherland, 2016; Bougher, 2017). From an epistemological standpoint, this raises the question: what is the essence of polarization—is it about beliefs, opinions, emotions, or trust?

The concept of affective polarization suggests that polarization can be analyzed as an emotional phenomenon. Such an analysis rests on three tenets: polarization is influenced by negative sentiments towards an opposing group, it involves passionately held beliefs, and it is connected to moral identity—a concept closely linked to emotions (Prinz, 2021). Studies on political emotions reveal distinct affective responses: conservatives tend to express fear and disgust, while liberals demonstrate empathy but also anger towards those they perceive as insufficiently empathetic (Inbar et al., 2012; Hasson et al., 2018; Yang et al., 2019). The applicability of these findings beyond the Democrat-Republican dichotomy remains uncertain. Still, the study of emotions reveals an initial critical distinction between individual and group emotions in the context of polarization.

Group emotions, stemming from the collective mood of a social group, can intensify polarization by driving group identity, solidarity within the group, and animosity towards outsiders (von Scheve, 2017; Versteegen, 2024; Renström et al., 2023). While individual-level and group-level emotions interact and reinforce each other, it is unclear how these emotional dynamics contribute to or reduce affective polarization. Emotions are dynamic, changing in strength and duration, and influenced by both individual and group emotion regulation strategies (Gross, 2015; Goldenberg, 2023). This suggests that polarization is a dynamic process rather than a fixed state, influencing the methodology for analyzing, measuring, and comprehending it. Affective polarization thus integrates emotional reactions at both the individual and group levels, prompting two issues to be briefly explored next: whether polarization is a process or a state and if it is an attribute of groups or individuals.

2.2 Across Multidimensional Spectrums: Interactions and Implications

Polarization manifests across a wide spectrum, encompassing ideological-affective, political-emotional, elite-mass, and individual-collective dimensions. In the United States,
for instance, it's clear that American political elites are both affectively and ideologically polarized. However, there is some debate as to the extent to which the American mass public is also ideologically divided or merely ideologically sorted (Fiorina and Abrams 2008). The American mass public is also clearly affectively polarized (Iyengar et al., 2012). How, if at all, do these dimensions interact?

The interplay between individual beliefs and systemic structures suggests that polarization is not just a matter of individual psychology or differing opinions, but it is also a state of the system that is reinforced by political, media, and social structures. It's a feedback loop where systemic incentives for division increase political homogeneity within groups, which, in turn, deepens systemic polarization (van Bavel and Pereira, 2018; Leonard et al., 2021; Levin et al., 2021; Callander and Carbajal, 2022; Simchon et al., 2022). In recent years, social media has been one determinant in this individual-collective nexus, as it progressively influences individual perceptions of the political landscape (Klein, 2020; Bail et al., 2018; Azzimonti and Fernandes, 2023, Törnberg, 2022). The proliferation of media channels and the advent of social media have created environments where individuals can be exposed almost exclusively to information that reinforces their existing beliefs and biases, thereby deepening convictions and increasing misinformation and distrust of the opposing side (Tucker et al., 2018; McLaughlin 2018; Kubin and von Sikorski, 2021).

Deeply interwoven into the individual-collective spectrum is the process-state spectrum, delineating the dynamics of polarization. The term 'polarization' may denote either a temporal snapshot (the configuration of a population at a given moment) or an evolving process (the dynamics of a population's configuration over time). Researchers typically regard polarization as static when analyzing opinion trends on assorted topics or within varied demographic cohorts. Conversely, they address polarization as a process when analyzing temporal shifts on a singular subject matter. Thus, when one observes that American political opinion is polarized, it is generally an articulation of a current, unchanging state; nonetheless, the term can also signify the process of becoming more polarized, such as the escalation of emotional polarization amidst electoral campaigns. Additionally, certain forms of polarization are inherently dynamic and cannot be captured through simple comparative analyses of groups or temporal sections. The interplay between intra-personal dynamics and inter-personal conflicts illustrate types of polarization that fall within a continuum from states to processes rather than fitting neatly at either extreme (van der Maas et al., 2020; Wu et al., 2022).

These nuanced variations and intersections across the diverse polarization spectrums are essential for comprehending the complex nature of polarization. It is not merely the interconnection and impact of various concepts and theoretical frameworks on polarization that are significant, but also the dynamic interplay between process-state and individual-collective spectrums, which influences the methods by which we quantify polarization and confront its ramifications. See Figure 1 for a visual representation.
Figure 1. Mapping polarization as a complex phenomenon. Navigating through the spectrum from individual to collective dimensions correlates with traversing the emotional-political and affective-ideological ranges. A selection of elements (e.g., beliefs, identity, trust) define polarization on an individual level and are situated within the individual/state quadrant. Progressing towards the collective/state quadrant, we encounter notions like social norms and collective emotions, which are pivotal for perceiving polarization as a trait of groups. The shift along the state-process spectrum may take place separately from the individual-collective transition, where factors like the influence of social media or stereotyping engage with the dynamic nature of polarization.

Understanding polarization as a process involves examining its genesis and possible effects. The origins of polarization have been studied from diverse perspectives, yielding varied and sometimes conflicting results. Classical social psychology research indicated that group discussions tend to amplify polarization, with individuals’ stances growing more entrenched as they assess the relevant evidence (Moscovici and Zavalloni, 1969; Lord et al., 1979). Contrarily, recent studies suggest that discussions could foster greater consensus among people, with the degree of social influence hinging on the preexisting similarity of their viewpoints (Axelrod, 1997; Hegselmann and Krause, 2002; Bramson et al., 2017; O’Connor and Weatherall, 2017). Alternate theories posit that polarization can stem from the strength of partisanship as a component of social identity or from the malleable nature of individual preferences shaped by the prominence of various information sources (Zaller, 1992; Iyengar et al., 2019). From these research angles, polarization is often regarded as a dangerous trend, potentially leading to the weakening of democratic principles, diminishing societal trust, and escalating levels of partisan antipathy and discrimination.

Nonetheless, empirical data suggests potential positive outcomes of affective polarization. Elevated polarization levels may indeed serve societal interests, correlating with increased political engagement and a broader sense of electoral options (Harteveld and Wagner, 2023).
Polarization can serve as a catalyst for mobilization, particularly engaging those typically disinterested in politics, thus resulting in higher voter turnout when oppositional feelings towards political adversaries intensify. Moreover, polarization may facilitate the decentralization of power, bolster collective decision-making accuracy, and amplify the voices of politically underrepresented individuals (Lupu, 2015; LeBas, 2018; Barrera-Lemarchand et al., 2023). If we opt for politics that reflect how we feel, political identities will inevitably diverge. Passionate polarization reflects the degree to which various groups have successfully brought their preferred ways of living to the forefront of public attention. Consequently, polarization reflects a societal preference for maintaining a mosaic of diverse identities over the acceptance of monolithic conformity (Hochschild, 2016; Prinz, 2021). These diverse impacts of polarization reinvigorate the debate over its nature and what precisely is being quantified when measuring polarization.

3. Measuring Polarization

Polarization, in its most fundamental sense, refers to the distribution of a particular construct across a population. For example, a population with polarized views on immigration would show a bimodal distribution of attitudes. However, polarization is more commonly understood as the division between groups within a system. Political polarization, specifically, denotes the divide between political parties and their supporters. The concept of polarization also captures divisions along various social dimensions, including class, race, geography, gender, and education. Definitions of polarization focusing on groups typically describe it as the divergence between groups or the consensus within them. A significant challenge for researchers examining countries outside the U.S. is the complexity of defining in-groups and out-groups in systems with multiple political parties.

Within the realm of political polarization, political polarization has been defined in many different ways with different target constructs, and debates over the extent to which political groups are polarized are, in part, a function of the target construct and the definition of polarization adopted. For instance, a central line of inquiry has been the extent to which Americans, en masse, are ideologically polarized (Abramowitz and Saunders, 2008; Fiorina and Abrams, 2008). Both sets of scholars rely on survey measures of policy attitudes, but different definitions yield different answers. One set defines polarization as ideological divergence, or movement to the extremes, interrogating whether the variance of attitudes within the mass public has increased over time (Fiorina and Abrams, 2008). Another set of scholars (e.g., Abramowitz and Saunders, 2008; Fiorina and Abrams, 2008) defines polarization as ideological coherence, wherein those who identify as Democrats increasingly hold liberal attitudes, while those who identify as Republicans increasingly hold conservative attitudes. Findings suggest that while ideological divergence has not notably increased, ideological coherence has intensified (Levendusky, 2009). Although the precise
impact of mass polarization on democratic stability is debatable, both divergence and coherence present challenges: divergence complicates consensus-building, and coherence fosters inflexible factions, undermining the potential for cross-party compromise.

Survey-based methods for studying political beliefs and ideological polarization are widely used but come with significant drawbacks. One issue is that people often don't answer these surveys sincerely (Bullock and Lenz, 2019). Some might choose extreme positions to show loyalty to a political party, while others might express opinions on policies they haven't actually thought about, resulting in what is known as "pseudo-attitudes." To address these limitations, researchers often turn to actual behaviors. For example, analyzing voting patterns is a common approach. When examining the general population, this often involves looking at which candidates people vote for. In studying political elites, researchers might analyze how legislators vote on specific bills, using this information to create a scale of ideological positions. Another innovative method involves assessing whom people follow on social media to deduce their political leanings.

Behavior-based methods, designed to address the sincerity issues associated with survey techniques, inherently rely on proxy measures to gauge political ideology, leading to debates over how accurately these behaviors reflect the underlying ideological positions. For example, the trend of narrowly decided elections favoring ideologically extreme candidates is often interpreted as a sign of widespread polarization. However, such outcomes might more accurately reflect the limited options available to voters rather than their actual ideological preferences (Fiorina and Abrams, 2008). Similarly, the degree to which roll-call measures of elite polarization or mass polarization using social media follower networks represent political beliefs is debatable. Neither roll-call votes nor online behavior may fully represent a person's ideology, in part because both are public and can be influenced by strategic considerations.

In addition to ideological polarization, researchers have begun to explore other types of polarization, with a particular focus on affective polarization. It is often measured using "feeling thermometers" on surveys, which assess an individual's likes or dislikes of a party. The prevalence of feeling thermometers owes more to their longstanding inclusion in large national surveys, facilitating longitudinal analysis than to their perfect encapsulation of affective polarization. Dislike for a political party can stem from various sources, including perceived social distance, identity affiliations, or ideological disagreements. Another critique of survey-based measures of affective polarization is that they typically ask respondents about their feelings towards, e.g., Democrats or Republicans (Klar et al., 2018). When answering these questions, respondents tend to think of political elites or political activists, which may yield results that overestimate the extent to which people dislike the average partisan.
The recognition of self-report limitations in accurately capturing affective polarization has motivated some scholars to turn toward behavioral indicators (e.g., Iyengar and Westwood 2015, 2017; Druckman et al., 2023; Dimant 2024; Dimant et al., 2024). These methods, which primarily assess trust, altruism, and cooperativeness, offer a novel lens to gauge partisan bias. Whether this bias is a result of partisanship of the partner, per se, versus other constructs that are strongly correlated with a partner's partisanship, e.g., their race or education, is an open question. At the forefront of this development are new validated measurement techniques that create the opportunity to measure important facets of polarization, including the ability to capture pluralistic societal views and norms (Dimant, 2023; Panizza et al., 2023, 2024). The broad implementation of such behaviorally informed scales will also enable researchers and practitioners to formulate policy implications that produce interventions that reduce polarization more successfully.

In sum, the diverse approaches to studying polarization—spanning ideological to affective dimensions—underscore the critical importance of ensuring that the chosen measures accurately reflect the targeted construct. Whether through surveys or behavioral indicators, researchers must navigate the challenges of capturing genuine attitudes and behaviors amidst the potential for insincerity, strategic responses, and the influence of external factors. Moreover, the reliance on proxy measures, while practical, introduces an additional layer of complexity in interpreting results and drawing conclusions about the underlying ideological or affective states.

3. Cause and effect in polarization research

At the heart of polarization research are assumptions about its deleterious effect on various social and political outcomes. Consequently, quite a bit of research (and money) is devoted to understanding its causes to identify interventions. Undoubtedly, various forms of polarization are negatively correlated with a wide range of outcomes, including democratic backsliding (Druckman et al., 2024). However, identifying whether these constructs directly influence other outcomes has proven to be a challenge. At the individual level, people who are, e.g., affectively polarized, also tend to be more ideologically polarized and more likely to be exposed to anti-democratic rhetoric. Several studies have manipulated (affective) polarization in the lab by, e.g., changing perceptions about the other side or strengthening real or imagined relations with the outparty (Broockman et al., 2022; Hartman et al., 2022; Voelkel et al., 2022; Dimant, 2024).

However, these experiments often violate the exclusion restriction, wherein the experimental manipulation also affects other related outcomes. For instance, playing a trust game with an outparty player moves inparty and outparty feelings in the same direction. A notable exception to this claim is Dimant (2024) showing with a series of experiments that ingroup-love and outgroup-hate can be varied independently. Additionally, existing studies
often rely on survey measures of affective polarization and survey measures of anti-democratic support and are susceptible to problems that apply to any self-reported survey measure. This research suggests that reducing affective polarization in the lab does not directly reduce support for anti-democratic outcomes outside the lab.

Finally, while researchers often presume that polarization will directly impact various outcomes, the specific mechanisms underpinning this relationship have not been well-explicated. Once these mechanisms are theorized, researchers and practitioners can begin the important work of ameliorating polarization. Additionally, even if these constructs don't directly impact various individual-level outcomes, they may interact with known causes of democratic backsliding, such as elite rhetoric.

4. Discussion and concluding remarks

Polarization has emerged as a significant factor influencing the fabric of democracy and societal cohesion. This phenomenon, intertwined with debates on climate change, the rise of populist movements, and responses to global health crises, serves as both a symptom and a catalyst of the broader democratic challenges we face today. The task of understanding polarization spans across disciplinary boundaries, necessitating an interdisciplinary approach to grasp its individual and systemic manifestations fully. This complexity is further compounded by the challenges in measuring polarization, with varied methodologies offering different insights and potential biases.

In this perspective piece, we discuss the various aspects of polarization through the interdisciplinary lens of behavioral science. Through this lens, our discussion centered around three main questions: What is polarization? How does it impact individuals and groups? And how can we measure it? Doing so has allowed us to explore its roots in identity conflicts and the erosion of trust in democratic processes. We have also highlighted the role of misconceptions in exacerbating polarization, which operates on both moral and psychological levels.

As we navigate the nuances of polarization and its implications, our discussion extends to the potential pathways for addressing and mitigating its divisive effects on society. We argue that polarization is a multifaceted phenomenon that extends beyond mere differences in opinion. It encompasses the ways in which these differences shape emotions, beliefs, and actions. Our discussion highlights the challenges associated with measuring polarization. Various methodologies exist, each offering different perspectives and drawbacks.

By addressing the critical questions of what polarization is, who it affects, and how it can be measured, we highlight the need for a comprehensive approach that embraces the complexity of this issue. The journey towards understanding and mitigating polarization is
fraught with challenges, particularly in developing measurement techniques that accurately reflect its diverse nature.

Looking ahead, the study of polarization is likely to evolve in several directions. First, there will be a greater emphasis on interdisciplinary research, combining insights from psychology, philosophy, political science, sociology, and data science. Second, the use of advanced computational methods and machine learning algorithms will become more prevalent in analyzing complex datasets related to polarization. Finally, there will be a growing focus on developing and testing interventions aimed at reducing polarization, with an emphasis on evidence-based approaches and real-world applicability.

From a broader perspective, it's worth exploring how polarization is interconnected with various other social phenomena. Viewed through a critical lens, an overemphasis on the analysis of polarization can sometimes overshadow other significant societal phenomena, such as extremism, prejudice, or group bias. While studying polarization sheds light on the division and discrepancies between groups, it may not necessarily reveal the inherent biases within these groups or indicate whether they are becoming more extreme. Extremism refers to ideologies, beliefs, or activities that deviate markedly from a society's accepted norms or dominant views. Much like polarization, extremism can vary significantly across different societies, cultures, and historical contexts. Key inquiries in the study of extremism may include determining the fundamental factors that cause individuals or collectives to adopt more radicalized stances or identifying effective methods for societies to deter and counteract extremism. Such questions explore facets of social conflict that, while related, are distinct from those addressed by polarization. Even though polarization garners significant academic focus, it should be considered one of the multiple interactive components within a broader system that are all vital to comprehensively understanding the intricate nature of social and political environments.

In sum, we advocate for a holistic approach to addressing polarization. It is essential to consider both the emotional and behavioral dimensions to develop a nuanced understanding. Such an understanding is vital for devising strategies that can mitigate the divisive effects of polarization and promote a more cohesive society. Doing so would enable both researchers and practitioners to better capture the richness of our societies and enable them to develop more effective solutions and policy implications.
Funding & Acknowledgments:
Kati Kish Bar-On was supported in part by the Institute for Humane Studies under Grant no. IHS017865. Yphtach Lelkes is a Fellow at Princeton Center for the Study of Democratic Politics.

References:


Electronic copy available at: https://ssrn.com/abstract=4784194
https://doi.org/10.1162/rest_a_01160.


Electronic copy available at: https://ssrn.com/abstract=4784194


Versteeghen, Peter Luca. 2024. “We Love, They Hate: Emotions in Affective Polarization and How Partisans May Use Them.” Political Psychology n/a (n/a). https://doi.org/10.1111/pops.12955.

Electronic copy available at: https://ssrn.com/abstract=4784194
