Misreading Social Identity Theory: Why Affective Polarisation Research Got It Wrong

Jingran Yang[*]

Abstract

This paper challenges the theoretical foundation on which contemporary research on affective polarisation rests. Since Iyengar et al. (2012), scholars have largely attributed partisan hostility to social identity theory, assuming that group membership automatically produces both in-group favouritism and out-group dislike. I argue that this interpretation is based on a fundamental misreading of Tajfel (1972)'s original theory. The minimal-group experiments that gave rise to social identity theory examined behavioural discrimination in resource allocation, not emotional hostility toward out-groups. By conflating discrimination with dislike, subsequent research has overstated the emotional consequences of partisan identity while neglecting the role of intergroup conflict. Drawing on Realistic Conflict Theory, this article develops a conflict-centred framework that distinguishes identity, discriminatory behaviour, and negative affect. It shows that hostile emotions towards out-groups arise from conflict rather than from social categorisation itself. This reinterpretation clarifies a long-standing conceptual error and situates affective polarisation within a broader theory of group conflict and political emotion.

1. Introduction

Affective polarisation—the growing dislike, anger, and hostility between political partisans—has become one of the defining features of contemporary democracies. Since the early 2010s, this phenomenon has been widely explained through the lens of social identity theory. In the most cited work, *"Affect, Not Ideology: A social identity perspective on polarization"* (Iyengar et al., 2012), the authors claim that partisan attachment operates like any social identity: it triggers positive feelings towards one's own party and negative evaluations of the opposing party. This interpretation has shaped an entire research agenda, linking partisan identity to emotions such as dislike, anger, and resentment. Yet this consensus rests on a serious theoretical misunderstanding.

A careful reading of Tajfel and Turner (1979) and related writings reveals that social identity theory was never intended to explain affective hostility. Their classic experiments measured intergroup discrimination in resource allocation, not hatred or dislike toward out-groups. The central finding—that individuals favour their own group even in minimal and meaningless settings—demonstrated that discrimination could exist without conflict or emotion. However, when later scholars translated this behavioural asymmetry into emotional animosity, they replaced *discrimination* with *dislike* and, in doing so, distorted the theoretical meaning of social identity.

This review thus makes two related interventions. First, it demonstrates that the current affective-polarisation literature misinterprets social identity theory by conflating neutral discrimination with emotional hostility. Second, it argues that this misinterpretation has diverted attention from an alternative and far more powerful explanatory framework—Realistic Conflict Theory (RCT). Developed by the Sherifs and later refined by Jackson (1993) and Sidanius and Pratto (1999), RCT explains how intergroup hostility arises from competition and threat over scarce resources, status, or power. Whereas social identity theory captures minimal differentiation without conflict, RCT captures the

2

very conditions—competition and threat—under which anger, resentment, and hostility emerge.

By reintroducing Realistic Conflict Theory into the study of affective polarisation, this paper reframes the problem from one of identity to one of group conflict. I argue that political partisanship alone cannot generate negative feelings towards outgroups, and that the hostility we observe in the electorate cannot be understood without separating group identity and conflict. In doing so, it not only clarifies a key theoretical confusion but also restores conflict—the very element Tajfel and Turner (1979) sought to isolate, not deny—to its rightful place in the study of political polarisation.

2. Social Identity Theory Does Not Predict Negative Out-Group Affect

Based on social identity theory (Tajfel, 1970; Tajfel and Turner, 1979), affective polarisation was first described by Iyengar et al. (2012) as partisan polarisation in order to differentiate it from policy-based divisions. In their most cited article, *"Affect, Not Ideology: A Social Identity Perspective on polarisation"*, the main idea of social identity theory was interpreted by them as stating that "any form of group membership, even one based on the most trivial of shared characteristics, triggers both positive feelings for the in-group and negative evaluations of the out-group." To match the empirical pattern of affective polarisation—where in-party feelings remained constant while out-party feelings have persistently declined since the 1970s—they further defined partisan polarisation as the extent to which partisans view each other as a disliked out-group.

However, this series of definitions by Iyengar et al. (2012) can be misleading. If readers carefully examine the original works of Tajfel and his co-authors, it becomes evident that Iyengar et al. (2012) mistakenly transformed *discrimination* into *dislike* when interpreting the effects of social identity as demonstrated in Tajfel's experimental studies on group behaviour. In the articles written by Tajfel (1970), Billig and Tajfel (1973), and Tajfel and Turner (1979), negative affections or feelings such as "dislike," "anger," or "hostility" toward out-groups—or positive emotions toward the in-group—

3

were never tested as outcomes of social identity. Instead, the only dependent variable in their experiments was discriminatory or prejudicial behaviour in resource allocation, rather than affective feelings toward out-groups or in-groups.

3.   Disentangle Discrimination and Negative Feelings

Discrimination and dislike are not interchangeable. As Tajfel (1970) himself wrote, "Discrimination, it is often said, is more directly a function of the objective social situation, which sometimes does and sometimes does not facilitate the expression of attitudes." Yet despite this clear distinction, the concepts of affection and discrimination have frequently been used interchangeably by later studies that cite this paper to explain negative affect toward out-groups (Iyengar, 2015; Iyengar et al., 2019; Druckman et al., 2018). This prompts me to suspect that many of those scholars who cited Tajfel (1970) may never have read his paper word by word. Instead, they might have relied solely on the interpretation of social identity theory provided by Iyengar et al. (2012); otherwise, such redinaccuracy could hardly have been made.

Before evaluating any relationship between negative affect and discriminatory behaviour, it is essential to distinguish the two, as they are conceptually distinct. Negative feelings toward out-groups—including dislike, hostility, anger, or prejudice—are fundamentally affective states that describe how individuals feel about members of other groups. Classic scholarship in social psychology has consistently emphasised this affective nature of intergroup evaluations. For example, Allport (1954) defined prejudice as a form of negative evaluation grounded in affect, illustrating the broader principle that emotional responses toward out-groups constitute attitudes rather than behaviours. Contemporary frameworks such as the SAGE Handbook of Prejudice, Stereotyping, and Discrimination (Dovidio, Hewstone, Glick & Esses, 2010) similarly distinguish affective reactions—whether prejudice, dislike, or hostility—from behavioural responses. Discrimination, by contrast, is explicitly behavioural. It refers to acting on group distinctions, typically through unequal treatment, differential resource allocation, exclusion, or preferential behaviour. Allport described discrimination as the "denial of equality of treat-

ment" based on group membership—a definition that highlights its behavioural nature. Modern research likewise conceptualises discrimination as observable actions rather than emotional states.

The tendency to treat negative affect as equivalent to discrimination likely stems from the fact that the two commonly co-occur. When negative feelings toward out-groups and discriminatory actions arise simultaneously, it is tempting to collapse them into a single construct. But co-occurrence does not entail conceptual equivalence. Negative feelings can be a sufficient condition for discriminatory behaviour in some cases, yet discrimination itself does not imply the presence of negative affect. This asymmetrical relationship between negative affect and discrimination follows directly from the fact that discrimination is a behavioural outcome. As a form of behaviour, discrimination can arise from a wide range of motivations, of which negative affect is only one possible source. Negative feelings—such as dislike or hostility—may lead individuals to allocate fewer resources to those they resent or to favour those they like, and such affect-driven discrimination is widely observable in everyday social contexts, including workplace evaluations, promotions, and interpersonal favouritism. In these cases, negative affect may be a sufficient condition for discriminatory action.[1]

However, because behavioural motivations are often multifaceted, the presence of discrimination does not imply that negative feelings are present. Discrimination can be entirely affect-neutral when it arises from motivations independent of emotion, such as instrumental rationality aimed at satisfying practical needs. Indeed, much of the discrimination observed in everyday life falls into this category. Widely accepted institutional practices—admitting high-performing students to better schools, awarding prizes only to top-ranked competitors, or offering different wages to workers with different skill levels— are all forms of differential treatment that do not stem from hostility toward any group.

---

[1]Some scholars even find little evidence that negative feelings reliably translate into discriminatory behaviour, suggesting that partisan animus may be neither sufficient nor necessary for discrimination to occur. For example, Lelkes and Westwood (2017) report no meaningful relationship between affective polarisation and intentionally harmful out-party discrimination: individuals at the highest levels of affective polarisation are no more willing to deliberately disadvantage the opposing party than those at the lowest levels, implying a near-independence between affect and discriminatory behaviour in their setting.

They function instead as mechanisms of incentive design, merit allocation, and efficiency enhancement. Economic examples make this logic even clearer. In market settings, price discrimination is a standard profit-maximizing strategy in which sellers charge different prices to different consumers to extract surplus. This form of discrimination conveys no animus; it reflects market segmentation and rational optimization. It is a paradigmatic case of discrimination without hostility—behavioural differentiation grounded in incentives rather than affect.

Discrimination may also arise from the self-serving nature of human behaviour rather than from negative feelings. Everyday resource-allocation decisions provide a clear illustration. Most individuals devote the overwhelming majority of their income to their families while donating only a small fraction to people outside their immediate circle. This pattern of differential treatment does not imply hostility toward out-groups. On the contrary, the very act of donating—even when individuals have full discretion to donate nothing—typically reflects positive regard or, at minimum, the absence of negative emotions. The reason most resources flow to the in-group is that self-interest naturally extends to those whose welfare is intertwined with one's own. In-group favouritism in this context is therefore an expression of self-regard generalized to the collective, rather than animus toward others.

Self-serving nature may also explain the mechanism in Tajfel's minimal group experiments. Once individuals are assigned to a group, even on the basis of trivial distinctions, they quickly begin to allocate resources in a way that favours their own group. This behaviour is not driven by dislike of the out-group—participants have no reason or opportunity to develop such feelings—but by the psychological extension of individual self-interest to the newly created in-group. What appears as intergroup discrimination is thus a form of non-hostile discrimination: behavioural differentiation rooted in self-regarding motivations rather than negative affect.

4. Group Conflict Theory: A Missing Foundation for Understanding Affective Polarisation

4.1. Partisan Identity Entangled with Conflict in the Literature on Affective Polarisation

If social identity cannot, by itself, lead to negative attitudes or feelings toward out-groups but only to discrimination, why do many later scholars find empirical evidence supporting a correlation between partisan identity and anger, dislike, or hostility toward out-groups (Valentino et al., 2011; Huddy, 2013; Groenendyk and Banks, 2014; Mason, 2015; Huddy et al., 2015a)? For instance, Groenendyk and Banks (2014) found that anger and enthusiasm, motivated by partisan identity, drive political participation. Mason (2015) demonstrated that the alignment of partisan and ideological identities generates anger toward partisan opponents more strongly than do simple issue disagreements. By conducting a series of experiments, Huddy et al. (2015b) further underscored the power of an expressive partisan identity to drive campaign involvement and to elicit strong emotional reactions to ongoing campaign events.

When Tajfel and Turner (1979) designed their experiments to demonstrate the effects of social identity, intergroup conflict or threat was intentionally redminimised so that their results would remain "relatively clean." After Iyengar et al. (2012), however, almost every study claiming that partisan identity drives anger toward partisan opponents or out-groups suffers from an impurity in research design: these studies were conducted in contexts where intergroup conflict or threat could not be eliminated. As a result, it becomes extremely difficult to discern whether the observed anger originates from partisan identity itself or from perceptions of intergroup threat.

In Groenendyk and Banks (2014)'s study, stronger partisans were found to express more anger than weaker partisans when they were primed with a threatening stimulus at the group level. What remains uncertain in their findings is whether, in the absence of any threatening stimulus, stronger partisans would still display anger toward the out-party. Similarly, in Mason (2015)'s research, although he concluded that the alignment of

partisan and ideological identities drives anger toward partisan opponents, he explicitly noted in a footnote that threat could also be a reason for such anger: "Though the out-group presidential candidate is not an ideal representation of threat, this candidate does represent the embodiment of the possibility that the in-group will lose status. If the out-group candidate succeeds, the in-group will have suffered a public failure."

Huddy et al. (2015a) likewise found that strongly identified partisans feel angrier than weaker partisans when they are threatened with electoral loss, and more positive when reassured of victory. Yet an important empirical finding in their paper—largely overlooked in subsequent literature—shows that threat, rather than party strength, exerts the more significant influence on anger toward the out-party. Because no existing research has successfully disentangled the effects of partisan identity from those of intergroup conflict or threat on negative feelings toward out-groups, it is premature to conclude that affective polarization originates from partisan identity.

4.2. The Group Conflict Theory: A Foundational Perspective Left in the Shadows

The Realistic Group Conflict Theory (R.C.T.) (Jackson, 1993; Baumeister and Vohs, 2007) is a social-psychological model of intergroup conflict (Sidanius and Pratto, 1999). However, it receives far less attention in the current literature on affective polarization than do social identity theory or group ideology theory. The theory was pioneered in social psychology by the Sherifs, who provided both an etiology of intergroup hostility and a theory of competition as *realistic* and *instrumental* in character—motivated by rewards that are, in principle, extrinsic to the intergroup situation (Deutsch, 1949; Julian, 1968). The theory explains how intergroup hostility can arise from conflicting goals and competition over limited resources, and it also accounts for the feelings of prejudice and discrimination toward out-groups that accompany intergroup hostility[2] (Jackson, 1993; Sidanius and Pratto, 1999; Whitley and Kite, 2010).

---

[2]Different from the social identity theory (Tajfel 1970) that inserts minimal intergroup discrimination can occur purely because of the group membership, no matter the intergroup conflict, the R.C.T. takes the development of in-group identifications and then intergroup discrimination almost as an epiphenomenon of intergroup conflict.

Groups may compete over real or perceived scarcities of resources such as money, political power, military protection, or social status (Jackson, 1993). Feelings of resentment can emerge when groups perceive such competition as a *zero-sum* situation—where only one group can win (obtaining the needed or desired resources) while the other loses (failing to obtain them because the "winning" group secures the limited resources first) (Jackson, 1993; Baumeister and Vohs, 2007). The length and severity of intergroup conflict, accordingly, depend on the perceived value and scarcity of the contested resource (Julian, 1968).

Although Tajfel and Turner (1979) pointed out certain inadequacies of R.C.T., their primary purpose in writing *An Integrative Theory of Intergroup Conflict*—where social identity theory was introduced—was, in their own words, "not to replace the R.C.T., but to supplement it in some respects that seem to us essential for an adequate social psychology of intergroup conflict." For unknown reasons, however, this original purpose was overlooked by the core literature on affective polarization. This oversight has led to an insufficient focus on Realistic Group Conflict Theory and, consequently, to a biased understanding of the roots of affective polarization.

Both social identity theory and realistic conflict theory examine intergroup behaviour and attitudes, yet they focus on distinct mechanisms. Social identity theory, as proposed by Tajfel (1970) and Tajfel and Turner (1979), emphasises the *pure label effect* of group membership—that is, the idea that "the mere perception of belonging to two distinct groups—that is, social categorization per se—is sufficient to trigger intergroup discrimination favouring the in-group." In other words, simply being aware of an out-group can provoke competitive or discriminatory responses from the in-group. Historically, this formulation was developed to challenge prevailing assumptions that intergroup discrimination arises primarily from ideological disagreement or intergroup conflict. Tajfel and his co-authors aimed to demonstrate that discrimination can exist even in the absence of ideological differences or overt conflict.

However, the contemporary literature on affective polarization, while attempting to separate social identity's pure label effect from group ideology, remains entangled with

the influence of intergroup conflict and threat. As a result, it fails to provide a comprehensive understanding of affective polarization. To achieve a fuller and more accurate explanation, scholars should analytically separate the effects of intergroup conflict and threat and reinterpret the phenomenon through the framework of Realistic Group Conflict Theory. redDoing so does not abandon insights from social identity or ideological theories, but rather clarifies the conditions under which they translate into affective polarization.

4.3. Group Conflict: Reconciling Inconsistent Explanations of Affective Polarization

A growing body of recent work suggests that affective polarizatio n declines with reductions in perceived intergroup conflict. For instance, Levendusky (2018) shows that strengthening American national identity can mitigate affective polarization: by making a superordinate national identity salient, the intervention dampens perceived partisan conflict and redirects attention away from party-based divisions toward shared national interests, thereby highlighting the potential for cooperation within a common ingroup. Related work points to a similar mechanism through cooperation. Lees and Cikara (2020), for example, argue that emphasizing intergroup cooperation lowers perceived competition intensity—and thus intergroup conflict—which in turn reduces outgroup animosity. Consistent with this conflict-centred perspective, Banda and Cluverius (2018) find that affective polarization increases as elites become more polarised, an elite-level pattern plausibly reflecting intensified issue- and policy-based conflict between the parties. Taken together, these findings suggest that intergroup conflict may be an overlooked yet pivotal driver of affective polarization, and that foregrounding conflict dynamics can help reconcile ongoing debates over its sources.

Although the literature on affective polarization has grown increasingly rich, existing explanations often appear fragmented. Scholars have attributed affective polarization to partisan identity, ideological polarization, issue-based distance, social sorting, elite competition, and individual-level characteristics, among others. Much of the debate—

especially between social identity–based and ideological accounts—has focused on identifying which of these factors constitutes the primary source of affective polarization. This apparent disagreement, however, may be overstated. From a mechanistic perspective, these explanations are not inherently incompatible. Rather, they converge on a common underlying process: affective polarization emerges when political differences are translated into intergroup conflict. Ideological polarization, for instance, does not uniformly generate affective polarization; only those ideological divisions that produce perceived threats, zero-sum competition, or group-based antagonism are likely to do so. Similarly, partisanship itself has not been shown to mechanically induce discriminatory or hostile affect, but partisan competition—particularly in electoral contexts—readily activates intergroup conflict. Social sorting operates through a similar channel by importing preexisting conflicts from nonpartisan identities into partisan alignments. Seen from this perspective, the diverse literatures on affective polarization are best understood not as competing explanations, but as alternative pathways through which intergroup conflict is generated and intensified. This conflict-centred view helps reconcile longstanding debates in the literature and suggests that intergroup conflict constitutes a core, yet under-theorized mechanism underlying affective polarization.

5. Emotions, Identity, and Outgroup Hostility Beyond Electoral Politics

Although the argument developed in this article is motivated by debates in electoral politics in the American context, its core mechanism is not specific to elections or to partisanship. As Weiner and Tatum (2021) argue, identity should not be treated as a unitary or self-explanatory variable, but as a political construct whose effects depend on processes of visibility, conceptualization, and recognition. From this perspective, identity provides a classificatory boundary, while its political and emotional consequences depend on the conflictual contexts in which it is activated. Building on this insight, the framework advanced here distinguishes between negative affect, intergroup conflict, and discriminatory behaviour, and shows how emotions toward out-groups emerge when po-

litical or social differences are translated into perceived group-level threats and zero-sum conflict. This distinction extends beyond electoral settings.

Evidence from research on violent extremism and radicalization illustrates the same mechanism. Pearlman's (2011, 2013) work on political mobilization in the Middle East shows that collective emotions—fear, anger, and hope—operate as proximate drivers of political action, while shared identity and structural grievances function as background conditions rather than direct causal forces. Even where collective identities remain stable, changes in emotional dynamics alter patterns of mobilization, demobilization, and violence. Evidence from experimental and neurocognitive studies also shows that hostility toward out-groups is not driven by ideological commitment or group identity alone; instead, perceived existential threat, group competition, and conflict escalation generate emotionally charged antagonism ((Pretus et al., 2018)). These affective responses arise from experiences of threat and exclusion rather than from prior identity transformation.

By clarifying the distinction between affective hostility and discriminatory action, and by centring intergroup conflict as the generative mechanism linking identity, emotion, and political outcomes, this framework offers a unified perspective that applies across both electoral and non-electoral domains. In doing so, it extends the study of affective polarization beyond American partisan politics and situates it within a broader theory of group conflict and political emotion.

6. Theoretical Stakes and Proposed Alternatives

The intervention advanced in this article carries important theoretical stakes for how political science conceptualises affective polarization and its social consequences. Much of the existing literature explains negative feelings toward out-groups by attributing them to social identity itself, often treating partisan, ideological, or group identities as the primary sources of hostility. Within this framework, negative affect is understood as a direct expression of identity-based attachment, and discriminatory behaviour is frequently taken as evidence of affective polarization.

This identity-centred interpretation is theoretically problematic in two important re-

spects. First, by locating the source of hostility in group identity per se, it risks essentializing negative emotions—implicitly suggesting that individuals dislike or resent others because of who they are, rather than because of any concrete conflict of interests, threats, or competition between groups. Analytically, this logic mirrors the cognitive structure underlying racial or religious prejudice, where hostility is attributed to immutable group characteristics rather than to situational or structural conditions. As a result, such explanations may inadvertently legitimise or reinforce the very forms of group-based animosity they seek to understand.

Second, treating affective polarization as an identity-driven phenomenon renders it analytically—and normatively—unsolvable. Social identities are an inherent feature of pluralistic societies and cannot be eliminated without eliminating social diversity itself. If negative affect toward out-groups is assumed to flow naturally from identity differences, then affective polarization appears as an inevitable and permanent condition of democratic life. This conclusion forecloses the possibility of mitigation and obscures the role of political institutions, negotiation, and conflict management.

By contrast, this article advances a conflict-centred alternative. It argues that negative feelings toward out-groups arise not from identity itself, but from intergroup conflict—including perceived zero-sum competition, threats to status or resources, and institutionalised rivalry between groups. Under this view, affective polarization is not an inherent feature of social identity but a contingent outcome of conflictual conditions. Importantly, conflicts are identifiable, negotiable, and subject to institutional regulation. This reframing restores analytical leverage and reopens the possibility that political systems can manage, channel, or mitigate negative emotions through conflict resolution, bargaining, and institutional design.

The theoretical stake, therefore, is not merely conceptual clarity, but diagnostic accuracy. Misattributing affective polarization to identity risks portraying democratic conflict as pathological and intractable. Recognizing intergroup conflict as the generative mechanism instead allows political science to distinguish between emotional hostility as a symptom and conflict as its source, thereby offering a more accurate—and more

hopeful—account of polarization in pluralistic societies.

7. Conclusion

This article revisits one of the most influential assumptions in the study of affective polarisation: that negative feelings toward political out-groups are a direct consequence of social identity. By returning to the original formulations of social identity theory and carefully distinguishing discrimination from emotional hostility, it shows that this assumption rests on a fundamental theoretical misinterpretation. Social identity theory was designed to explain minimal differentiation in behaviour, not the emergence of anger, resentment, or hatred. Treating it as a theory of affective hostility has therefore obscured, rather than clarified, the origins of contemporary political animosity.

Reintroducing Realistic Conflict Theory provides a more accurate and coherent account. Negative feelings toward out-groups arise not from identity itself, but from intergroup conflict—whether rooted in competition over political power, perceived threats to status, or zero-sum struggles over scarce resources. From this perspective, affective polarisation is a downstream outcome of conflictual conditions, not an inherent feature of pluralistic identities. Identity matters, but only insofar as it structures and channels conflict.

This reframing carries important implications for political science. By shifting attention from identity to conflict, it resolves longstanding debates over the sources of affective polarisation and avoids the theoretical and normative pitfalls of identity determinism. More importantly, it restores analytical leverage. If hostility is generated by conflict rather than by identity itself, then affective polarisation is neither inevitable nor irreversible. It becomes a contingent phenomenon—one that can be managed, mitigated, or intensified through institutional design, political competition, and conflict regulation.

Understanding affective polarisation in this way does not deny the depth of contemporary political divisions. It does, however, reject the conclusion that such divisions reflect an inescapable pathology of social identity. By locating negative affect where it belongs—in the dynamics of intergroup conflict—this article offers a more precise, more realistic,

and ultimately more constructive foundation for the study of political polarisation in democratic societies.

References

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

Baumeister, R. and Vohs, K. (2007). *Realistic Group Conflict Theory*, volume 2.

Billig, M. and Tajfel, H. (1973). Social categorization and similarity in intergroup behaviour. *European Journal of Social Psychology*, 3(1):27–52.

Deutsch, M. (1949). A theory of cooperation and competition. *Human Relations*, 2(2):129–152.

Druckman, J., Levendusky, M., and McLain, A. (2018). No need to watch: how the effects of partisan media can spread via interpersonal discussions. *American Journal of Political Science*, 62(1):99–112.

Groenendyk, E. and Banks, A. (2014). Emotions, ideology, and participation in politics. *Political Psychology*, 35(4):577–593.

Huddy, L. (2013). Group identity and political cohesion. In Huddy, L., Sears, D. O., and Levy, J. S., editors, *The Oxford Handbook of Political Psychology*, pages 737–773. Oxford University Press, Oxford, 2 edition.

Huddy, L., Mason, L., and Aarøe, L. (2015a). Expressive partisanship: Campaign involvement, political emotion, and partisan identity. *American Political Science Review*, 109(1):1–17.

Huddy, L., Mason, L., and Aarøe, L. (2015b). Expressive partisanship: Campaign involvement, political emotion, and partisan identity. *American Political Science Review*, 109(1):1–17.

Iyengar, S. (2015). Fear and loathing across party lines: New evidence on group polarization. In Norris, P. and Dalton, R. J., editors, *The Handbook of Electoral Behavior*, pages 284–305. Oxford University Press, Oxford.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., and Westwood, S. J. (2019). The origins and consequences of affective polarization in the united states. *Annual Review of Political Science*, 22(1):129–146.

Iyengar, S., Sood, G., and Lelkes, Y. (2012). Affect, not ideology: A social identity perspective on polarization. *Public Opinion Quarterly*, 76(3):405–431.

Jackson, J. W. (1993). Realistic group conflict theory: A review and evaluation. *The Psychological Record*, 43(3):395–415.

Julian, J. W. (1968). Processes of conflict resolution in experimental laboratory studies. In Pruitt, D. G. and Kimmel, P., editors, *The Structure of Conflict*, pages 211–233. Academic Press, New York.

Lees, J. and Cikara, M. (2020). Inaccurate group meta-perceptions drive negative outgroup attributions in competitive contexts. *Nature human behaviour*, 4(3):279–286.

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

Mason, L. (2015). "i disrespectfully agree": The differential effects of partisan and ideological sorting on affective polarization. *American Journal of Political Science*, 59(1):128–145.

Pearlman, W. (2011). *Violence, Nonviolence, and the Palestinian National Movement.* Cambridge University Press.

Pearlman, W. (2013). Emotions and the microfoundations of the arab uprisings. *Perspectives on Politics*, 11(2):387–409.

Pretus, C., Hamid, N., Sheikh, H., Ginges, J., Tobeña, A., Davis, R., Vilarroya, O., and Atran, S. (2018). Neural and behavioral correlates of sacred values and vulnerability to violent extremism. *Frontiers in psychology*, 9:2462.

Sidanius, J. and Pratto, F. (1999). *Social Dominance: An Intergroup Theory of Social Hierarchy and Oppression.* Cambridge University Press.

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, 223(5):96–102.

Tajfel, H. (1972). Experiments in a vacuum. In Israel, J. and Tajfel, H., editors, *The Context of Social Psychology: A Critical Assessment*, pages 69–119. Academic Press, London.

Tajfel, H. and Turner, J. C. (1979). An integrative theory of intergroup conflict. In Austin, W. G. and Worchel, S., editors, *The Social Psychology of Intergroup Relations*, pages 33–47. Brooks/Cole, Monterey, CA.

Valentino, N. A., Brader, T., and Jardina, A. E. (2011). Election night's alright for fighting: The role of emotions in political participation. *Political Psychology*, 32(2):187–211.

Weiner, S. and Tatum, D. S. (2021). Rethinking identity in political science. *Political Studies Review*, 19(3):464–481.

Whitley, B. and Kite, M. (2010). *The Psychology of Prejudice and Discrimination.* Wadsworth.