Northwestern | INSTITUTE FOR POLICY RESEARCH

Working Paper Series

WP-20-17

How Do Partisans Navigate Elite Intra-group Dissent? Leadership, Partisanship, and the Limits of Democratic Accountability

Alexandra Filindra
Associate Professor of Political Science
University of Illinois at Chicago

Laurel Harbridge-Yong
Associate Professor of Political Science and IPR Fellow
Northwestern University

Version: May 6, 2020

DRAFT

Please do not quote or distribute without permission.

ABSTRACT

Democratic erosion has led scholars to query how voters respond to leaders who violate norms. Given polarization and the centrality of identity in partisan affiliations, criticism by co-partisan elites may be crucial to checking party leaders. The researchers draw on theories of partisanship as a social identity as well as perspectives on leadership and dissent to theorize how partisans respond to misbehavior by an ingroup leader, and to criticism of the leader by a co-partisan. They test their expectations through multiple survey experiments. They find evidence of ingroup bias in evaluations of the misbehaving leader and little evidence that ingroup dissent is an effective constraint on leaders. Except in the most serious leadership transgressions of 'hard' norms, people rally around leaders when confronted with dissent by co-partisan elites. Overall, the results suggest that ingroup dissent may not lead to leader accountability.

Since his inauguration, President Trump has ignited numerous controversies that many view as violating democratic norms.[1] Democrats and even some Republicans have criticized Trump's actions. Yet, despite critics, the vast majority of Republican voters continue to support Trump. A January 2017 Gallup poll showed 89% of Republicans approved of the job Trump was doing as President; by February 2020, after his impeachment trial, the proportion remained unchanged (93%). While it is not surprising that Republican voters are not persuaded by criticism from Democrats (e.g., Iyengar and Westwood 2015, Huddy, Mason, and Aaroe 2015), it is more surprising that Trump seems immune to criticism emanating from within the party.

Donald Trump is neither the first nor the only contemporary leader to violate democratic norms. Issues of political accountability have emerged in Hungary, Poland, Brazil, Israel, Turkey, and elsewhere. The continuing support that leaders receive from co-partisan voters despite fervent criticism from domestic and external sources has contributed to the growing literature on democratic backsliding. Yet, much of this literature, mostly situated in Comparative Politics, focuses on elite competition and the role of norms and institutions as constraints on transgressive elite behavior (Levitsky and Ziblatt 2017, cf., Foa and Mounk 2016, Svolik 2019). By contrast, we focus on American public opinion and ask whether, in the context of hyperpolarization, the co-partisan demos are likely to penalize leaders who may be a threat to democratic institutions, particularly if criticism emerges from within the party.

These leaders' apparent "Teflon" quality with the party base presents a dilemma for political accountability. Many models of voter decision-making that emphasize the role of popular sovereignty assume that voters intrinsically appreciate democratic norms (Jennings and Niemi 1974), and that when voters are exposed to negative information about leaders, especially when contemptible actions are visible and frequent, partisans should update their views downward, thus holding leaders accountable (e.g., Fiorina 1981). Although the public appears to understand when violations of democracy occur (Carey, Clayton, et al.

[1] For example, CNN's list: https://www.cnn.com/2013/07/04/us/donald-trump-fast-facts/index.html.

2019), these "hard" assumptions of democratic accountability have been questioned by several strands of research: first, voters lack ideological constraint (Abramowitz and Saunders 2008); second, people make decisions based on heuristics (Zaller 1992); and finally, partisanship is rooted in group identity making ingroup members susceptible to ingroup favorability bias and outgroup antagonism (e.g., Mason 2018, Barber and Pope 2019, Iyengar and Westwood 2015), which in return means that partisans are less willing to penalize transgressions by ingroup leaders relative to partisan outgroup (Carey, Clayton, et al. 2019, Svolik 2019). These latter two points suggest that criticism of leaders by co-partisan elites might be the most likely case for shifting partisan opinion as outgroup elites carry little credibility among ingroup members. From this perspective, public constraints on partisan leaders may rest on whether co-partisan elites voice criticism when party leaders violate democratic norms.

In a recent book, Levitsky and Ziblatt (2017) evoke this expectation by arguing that Republican voters will abandon Trump only if other Republican leaders—i.e., trusted ingroup elites—consistently raise their voices against the President. However, there is no evidence theoretically or empirically that this is a valid prediction. In fact, in a recent study, Walter and Redlawsk (2019) show that people appear resistant to punish ingroup leaders even when the leader's behavior goes against partisans' moral. .Rationalistic theories of political evaluation, which emphasize elite performance in domains such as the economy, effective governance, and combating corruption, suggest the public has expectations of political leaders and conceptualizes political support as a "running tally" of retrospective evaluations (e.g., Fiorina 1981, Mishler and Rose 2001). Since corrupt political behavior is a violation of norms, if not of laws, such theories suggest that voters should take such behavior into account and penalize leaders who violate the norms of democracy. However, the little evidence we have suggests that this may not be the case in the context of hyperpolarization (Carey, Clayton, et al. 2019, Svolik 2019).

In this study, our key questions are: 1) do co-partisans recognize and penalize leader misbehavior on their own? 2) even if co-partisans don't make the connection unassisted, when faced with credible *ingroup*

dissent to an ingroup leader's misbehavior,[2] do partisans accept the criticism, decreasing their support for the leader, thus increasing the likelihood of political accountability, or do they double-down?

We draw on partisan social identity theories in political science, as well as psychological theories of group leadership and dissent, to theorize how partisans may approach norms-violating leaders and co-partisan dissenters who criticize such norms violations (e.g., Hogg 2001, Jetten and Horsney 2014, Steffens, Haslam, and Reicher 2014). Not only are the answers to these questions important in their own right, but they shed light on broader debates in democratic politics. While elite perspectives on democracy (Levitsky and Ziblatt 2017) and scholarship on good government focus on institutional checks (Rothstein and Varraich 2017), our research points to the potential (and limits) of punishment by the co-partisan mass public. Likewise, scholarship on party competition suggests that robust partisanship can be valuable for political engagement and democratic responsiveness (e.g., Rosenblum 2008), but our research points to ways that partisanship may *reduce* the accountability of misbehaving elected officials.

We explore this perspective by leveraging three online survey experiments. Our results suggest that partisanship changes the way people evaluate political misbehavior. Misbehaving leaders benefit from ingroup favoritism, especially when voters have actual affective ties to them. Ingroup dissent tends to have the opposite effect from what perspectives focused on instrumental support for political institutions suggest: there is an affective "rally around the leader" effect which, for real leaders, could extend to boosting support in a primary election. Lastly, while dissent is typically viewed as appropriate, it is not generally rewarded,

[2] In this article, we use the term "misbehavior" to describe actions that violate democratic norms in some way. These types of norms are unrelated to liberal or conservative ideological perspectives. They include: free and fair elections, judicial independence, and no abuse of office (for other examples see http://brightlinewatch.org/us-elections/, Levitsky and Ziblatt (2017)). Here, we do not engage with violations of personal norms, for example an extramarital affair.

except for cases of serious norms violations. Taken together, our results show that holding misbehaving partisan leaders accountable for their actions is a major challenge for which co-partisan dissent is not panacea.

Theoretic Framework

Social Identities and Ingroup Political Leadership

According to Social Identity Theory (SIT), people have a predisposition to favor their ingroup. Preserving group identity boundaries is a key goal, whether for individual validation or due to psychological needs to maintain firm intergroup boundaries (Steffens et al. 2015). Even when it comes to responding to transgressive leaders, people are not blind but rather "engaged" followers willing to obey authority when they identify with the leader (Reicher and Haslam 2011, Reicher, Haslam, and Rath 2008).

Research on partisanship and polarization indirectly confirms that people do not follow all authority figures, all of the time, but defer to ingroup leaders. Specifically, because partisans operate as members of a group, when confronted with interparty conflicts they rely on group biases (Taber and Lodge 2006), and are skeptical of leaders of the out-party (Mason 2018, Nicholson 2012). Most recently, scholars show that partisans may condemn ethical lapses or democratic norm violations of the out-party but not their own (Walter and Redlawsk 2019, Carey, Clayton, et al. 2019, Graham and Svolik 2019). This suggests that people may defer to ingroup leaders even when leaders violate norms.

Beyond the pull of the ingroup, the hierarchical position of the leader may also contribute to their ability to violate norms without consequence. Psychology shows that leaders play an important role in groups because they are perceived to combine the key positive attributes that members associate with their group (Steffens et al. 2015, Hogg 2001). The more representative of these qualities a leader is perceived to be, the greater the support she receives from the ingroup. Thus, leaders become the yardstick against which group members determine their own standing in the group and the normative distance from outgroups. Thus, there is an incentive for group members to identify strongly with *group leaders* as this ensures maximum

psychological distance from outgroups (Steffens et al. 2015). The result is that group members are loyal to group leaders and perceive them as representative of the group identity (Giessner, van Knippenberg, and Sleebos 2009), which, in turn, increases the strength of group attachment (Steffens, Haslam, and Reicher 2014, Steffens et al. 2015). Overall, the literature suggests a strong endogenous relationship where identifying with leaders generalizes into stronger bonds with the ingroup and stronger connections to the ingroup translate into a stronger attachment to leaders.

Since evaluations of leaders can translate into evaluations of the ingroup, these assessments are not tethered to objective measures of success and failure, or even moral foundations (Walter and Redlawsk 2019). Group members have incentives to downplay leader failure and over-emphasize success (cf., Rast et al. 2015) and group leaders may be granted a "license to fail" (Giessner, van Knippenberg, and Sleebos 2009). Followers are more likely to interpret questionable/unsuccessful leader actions from the prism of "she is doing it for us" and therefore approach the situation with charity (Haslam et al. 2001). Combined with insights on partisanship as identity (Mason 2018), this perspective suggests that the pull of ingroup party leaders may be stronger than partisanship alone, thus increasing co-partisan followers' incentives to rationalize/excuse their leader's misbehavior.

But what happens when prominent ingroup members, even speaking on behalf of the party, criticize the actions of leadership, highlighting normative violations? Although this has been proposed as a key check on partisan leadership (Levitsky and Ziblatt 2017), little is known about how partisans navigate ingroup conflict where, by definition, partisan cues have no heuristic value. How do individuals respond to leaders whose behavior is at odds with group standards, or to ingroup critics? Does the ingroup identity of the dissenter make him effective or does the centrality of the leader to group identity make her immune to criticism?

Social Identity and Ingroup Dissent

Understanding how group members approach intragroup conflict, and specifically dissent to leaders

misbehavior, requires a discussion of how ingroup members engage with two different types of norms violations—deviance and dissent. Whereas deviance is defined as the violation of group norms (e.g., democratic norms), dissent is defined as the act of disagreement with the group/leader (Jetten and Horsney 2014, 463). Even if dissent occurs with the best interests of the group in mind, it may not be seen positively by group members. This suggests that even if the norms-defying leader may be viewed as "deviant," the critic's dissent may also be seen as "deviant" since it goes against the group leader. Both may pose challenges for ingroup members who generally seek to protect the boundaries of group identity and may limit the effectiveness of critics at moving ingroup opinion of leaders.

If co-partisans seek to protect the boundaries of group identity, criticism of the leader can produce cognitive dissonance. Individuals may be motivated to reduce dissonance by justifying the behavior of leaders as necessary, suggesting ingroup criticism could lead not only to a lack of Bayesian updating, but to even greater rallying around leaders (e.g., Redlawsk 2002). In conjunction with the research discussed above that highlights the connection between leaders and group identities, this suggests that leaders may be insulated from criticism and dissent may not affect co-partisan voters' evaluations of them.

There are also several reasons that ingroup dissent may be seen as worse than the initial norms-defying behavior by the leader. Restoring the positive image of the group, protecting cohesion, and reinforcing boundaries are key motivations in determining responses to dissent (Hutchison et al. 2008). Ingroup challengers can be punished by members (Hogg 1992, Hornsey et al. 2006) and the "black sheep effect" highlights the importance of ingroup identity by suggesting that members are more punitive toward *lower* status ingroup norm violators rather than outgroup members committing the same offense (e.g., Marques, Yzerbyt, and Leyens 1988).[3] Prior commitment to leaders can also influence group members' response to

[3] This effect may be particularly strong when the dissent relates to a core dimension of group identity, when the group is under threat, or when the dissenters embarrass the group (Abrams et al. 2008, Branscombe et

dissent (Monin, Sawyer, and Marquez 2008). Given a leader's role in a social group, criticism of a leader can be interpreted as implicit criticism of the group itself and thus threatening. Combined with the perspectives on partisan social identity and leadership, research on dissent suggests that not only might criticism of party leaders be ineffective, but it may harm the dissenting member.

Research Design

Our aim is to assess whether, as perspectives on instrumental support for political institutions would suggest, partisan voters are critical of misbehaving leaders, particularly when other co-partisans dissent, or if dissent may not have the intended effect on partisans' perceptions of party leaders, as social identity theories imply. In the hypotheses we lay out below we contrast expectations from the instrumental support perspectives with expectations of an ingroup leadership bias. Specifically, the former suggests that people can rationally evaluate the behavior of leaders and outputs of institutions and can recognize and punish corruption and ineffective governance; thus the (co-partisan) public should effectively hold leaders accountable, especially when a trusted source provides damning information (Fiorina 1981, Hetherington 2005). However, the latter highlights the limits of accountability for ingroup leaders. Because our hypotheses are more easily understood when presented in direct connection to the experimental design, we first explain the study design and then we outline our expectations. In the studies that follow we also explore whether the limits of public accountability differ for leaders of a neutral outgroup (where, absent partisanship and affect for the leader, accountability may be most likely), a hypothetical ingroup leader (thus controlling for leader affect), or a real ingroup leader; comparisons that allow us to speak to the breadth of

---

al. 1993, Chekroun and Nugier 2011). Yet it is important to note that dissent is not always rejected and punished. Rather, response to dissent is conditional upon the dissenter's standing (Galinsky et al. 2008), perceived intentions (Hornsey, Trembath, and Gunthorpe 2004), and seriousness of criticism (Hornsey and Jetten 2003).

deference to leaders and whether ingroup leadership bias applies to both hypothetical and real leaders, or only to real leaders that people may clearly see as representing their group identity. Figure 1 summarizes the design for all three studies that we present. Details on sampling, sample sizes, data collection, and vignette wording are included in Appendix A.

[FIGURE 1-HERE]

In all three studies, respondents were assigned to either an ingroup condition (Republican or Democrat, matched to respondent partisanship) or a neutral control condition (Australia). In two of the three studies, respondents in the ingroup condition read a story about a fictitious state Governor (Provincial Premier); in the third study, which included only Republican respondents, a third condition referencing President Trump was included. In all studies, the story suggests that the leader violated democratic norms. After responding to feeling thermometers (affect) and assessments of how party members should vote in the next primary election (vote),[4] designed to gauge their initial view of the leader's behavior, respondents were randomly assigned to either a loyalty condition, where they read that the co-partisan leader in the state Assembly defended the Governor's (Premier's) action, or to a dissent condition, where the co-partisan Assembly leader criticized the Governor's (Premier's) action.[5] In the case of Trump, the dissenting ingroup

[4] With the exception of the Trump condition, we could not ask respondents about their own intent to vote because the state governor was not identified with a particular state and Americans cannot vote in Australian elections. However, we believe that their recommendation also represents their own intended action.

[5] Designating the co-partisan supporter or critic as the *leader* of the party in the legislature is done to emphasize the party's response to the leader and to minimize any confounding related to large status differentials; studies show that ingroup members are more likely to punish a marginal group member who is critical of the group than a core member (Jetten and Horsney 2014).

leader was Senate Majority Leader McConnell. Following this part of the story, respondents were asked to evaluate the critic (appropriateness of action, feeling thermometer, and recommendation for primary voting), and provide a new set of evaluations of the misbehaving leader, in light of this new information (feeling thermometer and vote).[6]

The studies vary in two key ways. First, they vary in whether hypothetical or real leaders are referenced. Studies 1 and 2 leverage hypothetical leaders, allowing us to isolate the impact of ingroup identities independent of previous affect toward a leader and allow us to examine both Democrats and Republicans. Study 3 focuses on Republicans, where we have the opportunity to compare responses to hypothetical ingroup leaders and a popular Republican leader: Donald Trump.

Second, the studies differ in the seriousness of the norms-defying behavior of the leader. We do this in two ways. First, the studies differ in whether the violation falls under "soft" or "hard" norms (Levitsky and Ziblatt 2017). Studies 1 and 3 focus on a legally-permissible but ethically objectionable pardon (soft norms), while Study 2 focuses on a politically-motivated investigation of judges, essentially an extra-judicial attempt to intimidate a co-equal branch of government. This may also be a crime, although it is not discussed as such in the story (hard norms). In Studies 1 and 3, subjects read a story describing the Governor's decision to pardon a party insider who had been convicted of violating campaign finance laws.[7] In Study 2 we increased the seriousness of the misbehavior by focusing on violations of hard norms. The story informed respondents

[6] To account for the possibility of anchoring effects, we designed two pilot tests in which the respondent is informed of both the co-partisan leader's transgression and the Assembly leader's response in the same text. Patterns of responses to the leader are similar to those reported here.

[7] Although this behavior would be considered "scandalous," this is not simply a political scandal as it has implications for the proper function of a democratic system. We designed our treatments after reviewing the list of democratic norms listed and benchmarked by Carey, Helmke, et al. (2019).

that in response to a state Supreme Court decision that went against him, the Governor ordered that law enforcement investigate the personal lives and finances of the judges who opposed him.[8] The Court is implicitly assumed to have operated within its bounds and made a decision on the merits of the case. As such, it brings us closer to behaviors that emerge in authoritarian regime contexts and which are (presumably) unacceptable in democratic politics. Both the pardon and the intimidation of judges are self-serving acts that constitute norms violations; however, one is legally proscribed, while the other is only normatively proscribed.

Second, within the "soft norms" violation (pardon), we consider two different contexts of the norms violation that may affect the seriousness of the leader's action. In Study 1, respondents were randomly assigned into one of two levels of seriousness. Either they were told that the party insider had acted illegally to benefit the party (low seriousness – because the act was out of party loyalty), or that the pardoned party insider had acted illegally to help the party but *also had ties to the Mafia and pocketed some of the ill-gotten contributions* (high seriousness – because the act involved self-dealing not just party loyalty). The Governor's behavior in both cases falls under the category of violations of "soft" norms because Governors have absolute authority to pardon. However, we considered that violations of soft norms may not be perceived in a binary fashion; the context may be important in partisan evaluations of leaders. Increasing the seriousness of the pardoned offense may affect perceptions of the seriousness of the norms violation.

In Study 3, which was fielded among Republicans only, we used the same low seriousness condition as in Study 1, but in addition to the neutral control and the partisan Governor condition, we included a condition that referenced President Trump to test the effect of norms violations and dissent with real leaders.[9]

---

[8] This is also selected from the list provided by Carey, Helmke, et al. (2019).

[9] We opted not to replicate this with Democratic respondents because we lack an equivalent to President Trump. Using a former President or an actual Democratic Governor could introduce confounders. First, a former leader and an extant leader are not psychological equivalents. Also, the most recent Democratic

Three additional aspects of the design are of note. First, we compare to a neutral rather than an opposing-partisan outgroup leader because studies of negative partisanship suggest that the level of negative affect towards out-partisans is very high in the United States (Iyengar, Sood, and Lelkes 2012). Previous studies have documented that in the context of hyperpolarization, partisans respond more critically when an outgroup member violates norms or engages in other disliked behaviors (Carey, Clayton, et al. 2019, Svolik 2019, Walter and Redlawsk 2019).Thus, our approach takes a more conservative test to identifying ingroup bias and separates it from the dislike of partisan outgroups. In the absence of outgroup bias, perspectives emphasizing instrumental support for institutions are most likely to hold: according to Carey, Clayton, et al. (2019) people know what democratic norms are and by extension, they should penalize violations when partisan identities do not enter their decision. Second, we opted to locate our control outgroup leader in Australia for several reasons. Australia is a federation bearing similarities in structure to the U.S. and the Premier is a real office, similar to a U.S. state Governor. Moreover, it is an English-speaking, mostly European-descending country. This allowed us to keep the names of our politicians consistent across treatments. In addition, both Democrats and Republicans have positive views of Australia and consider it a key ally, so their responses to the story are not likely to be colored by negative views of the country itself (Katz and Quealy 2017).[10] Finally, given the generally low political knowledge among Americans (Delli Carpini

President was Barack Obama which introduces the confounder of race. Bill Clinton, the previous Democratic President, ended his tenure in 2000 and many people, especially younger Democrats, may not even have experienced his presidency as adults. A governor is not the same political focal point as the president of the United States, so such a comparison to Trump may not be appropriate especially for those outside the governor's state.

[10] In a separate pre-test we found that while Democrats viewed Australia as more likely to do the right thing in its relationship with the U.S. than Republicans (Mean=0.790 for Democrats, Mean=0.687 for Republicans

and Keeter 1996), it is unlikely that our respondents know very much about politics in Australia.[11] For the Australian treatments, we used the "National Party" rather than the Conservative or Labor parties because Americans may have affective priors to those labels – which are similar to ideological and group identities of U.S. parties – which could impact the results.

Third, in the Democratic and Republican versions of the treatment we did not include a mention of the Governor's state. We did this to avoid confounding the results with respondents' knowledge of a given state's politics and their affect for that state. Finally, we used a fictitious name for the Governor (also, the Assembly leader and in Studies 1 and 3, the pardoned party insider), and one that evokes white, Anglo-Saxon males. We opted to do so in order to avoid any confounders arising from misperceptions of the characters' race and gender.

As a note on nomenclature, given the complexity of the experiments, we use the following terms in the analysis to describe various conditions: "misbehaving leader" refers generally to the norms violating leader in both the control and the ingroup conditions, "Australia" refers to the control condition of the neutral outgroup leader, "Governor" refers to the partisan ingroup leader who misbehaves, and "Assembly leader" refers to the political leader who responds to the misbehavior in both the control and the partisan conditions. For Studies 1 and 3, the "businessman" is the person who was pardoned.

Hypotheses

Given the structure of the experimental design, we focus on three sets of outcomes: evaluations of the misbehaving leader *before* respondents read about the co-partisan Assembly leader's response,

---

on a 0-1 scale, p=0.007), respondents from both parties view Australia more favorably than France, China, Russia, or Iran.

[11] Canada would have also met the criteria of a federal country with a European heritage, but it is more likely that respondents would be knowledgeable about Canadian politics.

evaluations of the misbehaving leader *after* the co-partisan Assembly leader's response, and evaluations of the co-partisan Assembly leader. For each set of outcomes, we lay out alternative hypotheses that follow from instrumental perspectives of democratic accountability as well as expectations that follow from the partisan social identity (Mason 2018) and ingroup leadership (Hogg 2001) perspectives discussed above. Whereas instrumental support theories suggests that both ingroup and neutral outgroup misbehaving leaders will be evaluated similarly and co-partisan dissent against them should be effective, the ingroup leadership bias hypotheses suggest that shared partisanship between respondents and the leaders, combined with the pull of leadership, will lead to decreased accountability.

Evaluations of the misbehaving leader before the Assembly leader response:

H1a (Instrumental support for institutions): Evaluations of the neutral outgroup misbehaving leader (control) = evaluations of hypothetical ingroup Governor = evaluations of real partisan leader

H1b (ingroup leadership bias): Evaluations of the neutral outgroup leader (control) < evaluations of the hypothetical ingroup Governor/evaluations of the real partisan leader

Evaluations of the misbehaving leader after the Assembly leader response:

H2a (Instrumental support for institutions): Across all groups, evaluations of the misbehaving leader in the loyalist condition are higher than evaluations of misbehaving leader in the dissent condition (because the new information allows people to update their evaluation of the leader).

H2b (ingroup leadership bias): In both the hypothetical ingroup misbehaving leader condition and the real leader condition, dissent is expected to be ineffective: it is either ignored or it produces a rally around the ingroup leader (evaluation of misbehaving leader in loyalist condition =< evaluation of misbehaving leader in dissent condition).

Evaluations of the co-partisan Assembly leader:

H3a (Instrumental support for institutions): Respondents reward the dissenter for calling out wrongdoing. Across all groups, evaluations of the loyalist Assembly leader are lower than evaluations of the dissenting

Assembly leader (as people reward the dissenter for providing new information relevant to a voter's political calculus and for taking the normatively appropriate position).

H3b (ingroup leadership bias): For both the hypothetical Governor condition and the real leader condition, partisans penalize the dissenter for being disloyal and threatening the integrity of the group.

Furthermore, the combination of studies in our design allows us to assess whether any ingroup leadership bias extends to *both* hypothetical and real leaders, or, whether it is confined to real leaders with whom respondents have affective group attachments. For instance, in hypothesis 2b, partisan affiliation may be necessary and sufficient to produce a threat to the integrity of the group and to produce a rally effect in response for both the hypothetical and real leaders. Alternatively, partisanship may be necessary but not sufficient to produce a threat to the integrity of the group and a corresponding rally effect; rather, it requires real ingroup leaders who are central to partisans' identities to produce a "rally around the leader" effect.

Data & Analysis

Study 1 was fielded with Lucid from Oct. 24-26, 2018 and included 1,379 respondents. ***Study 2*** was fielded with Lucid from May 29-31, 2019 and included 1,752 respondents. Both samples were balanced to national census demographics on gender, race, region, and age (prior to dropping independents).[12] ***Study 3*** was conducted on Amazon's MTurk from October 24-25, 2018 and included a total of 760 respondents who completed the entire questionnaire. Only respondents who self-identified as Republican were included in the study; independents and Democrats were screened out. Appendix Tables A1-A3 show the number of respondents for each condition in each of the three experiments. Balance tables and descriptive statistics for each of the studies are in Appendix A, Sections A3 and A4.

For ease of comparison, we present results from all three studies concurrently rather than

[12] For both studies, this number excludes pure independents, speeders and partial completes.

sequentially. All measures included in the analysis were re-coded on a 0-1 scale to allow comparability across measures. Figures start with Study 1 and proceed left to right to Study 3. The analysis presents pooled partisan results for several reasons. First, our theoretical expectations do not differentiate between parties and the literature on partisanship suggests symmetrical behaviors. Second, we did not manipulate partisanship in the experiments. Third, the results confirm no systematic and consistent differences in responses across partisan groups.[13]

Crime Seriousness - Studies 1 & 3

Before turning to the results, we briefly discuss whether respondents recognized the differences in seriousness of the pardoned crime in Studies 1 and 3. Respondents' evaluation of the underlying pardoned crime is an important component of the experiment because our theoretical framework focuses on leader behaviors that defy norms. Although the leader's behavior in Studies 1 and 3 is the same in that he pardons the businessman, when the underlying crime is more serious, the leader's response may be seen as a more alarming violation. For instance, if the criminal behavior is more egregious or self-serving, the pardon may be seen as occurring solely because of financial contributions to the party. For Studies 1 and 3, both our pre-treatment measures and a pre-test study suggest that campaign finance violations are viewed as a moderately serious crime.[14] First, in a pre-treatment question in Study 1 we asked partisans to rank six different crimes in terms of seriousness (re-coded 0-1 where higher values are more serious). Respondents

[13] This is consistent with other findings such as Carey, Clayton, et al. (2019) and Svolik (2019). These studies find symmetrical support for partisan candidates who violate democratic norms, but do not address questions of ingroup dissent.

[14] Extant research shows that the public takes campaign finance violations seriously and is willing to punish violators (Wood and Grose 2019).

ranked campaign finance violations as moderately serious (Mean=0.424).[15]

In a pre-test study, we also tested evaluations of the relative seriousness of our low seriousness crime (campaign finance violations) and high seriousness crime (campaign finance violations + self-dealing/mafia ties) conditions on a 5-point scale (recoded 0-1). We found no partisan difference in initial rankings of the crime seriousness relative to other white-collar crimes ($Mean_D$=0.393, $Mean_R$=0.332, p=0.149). Moreover, when given similar descriptions of the crime as those used in Study 1, respondents viewed the crime as more serious when the businessman was described as self-dealing ($Mean_{low}$=0.625, $Mean_{high}$=0.702, p=0.03). Adding information about the businessman having ties to the mafia also increased the perceived seriousness of the crime ($Mean_{low,\ no\ mafia}$=0.625, $Mean_{high,\ mafia}$=0.789, p<0.001). These patterns highlight that people not only recognize campaign violations as a crime but see it as a more serious crime when the convicted individual has a criminal background (Mafia ties) and personally benefits from the crime.

In Studies 1 and 3, after respondents were exposed to the pardon story, we again asked them to evaluate the seriousness of the crime being pardoned. Whether pooling the data or analyzing Democrats and Republicans separately, we see that respondents in Study 1, where crime seriousness varied, react to the crime seriousness conditions as expected, viewing the high seriousness crime as significantly more serious (Appendix A, Section A5).

Despite recognizing that the treatments varied in seriousness, we find that crime seriousness has no differential effects on the evaluation of the misbehaving leader or the Assembly leader. This is an important finding in itself because it suggests another dimension of the "principle/implementation" disjuncture, meaning that the way partisans evaluate a criminal offense in the abstract diverges substantially from the way they

[15] Democrats considered this a more serious crime than did Republicans ($Mean_D$=0.447, $Mean_R$=0.397, p=0.005) perhaps reflecting contemporary partisan discourse around these issues.

respond to political leaders whose actions (i.e., pardon and response to pardon) should be evaluated in terms of the seriousness of the original transgression. In this sense, people seem to understand the political context as separate from the criminal context.[16] Given the null findings on the effects of crime seriousness on misbehaving leader and Assembly leader evaluations, for the remaining analyses presented herein, we pool across the low and high seriousness conditions in Study 1, effectively analyzing a 2x3 experiment (Appendix A, Section A6).

Seriousness of Violation - Study 2

In Study 2 (investigation of judges), respondents are not asked to evaluate two separate behaviors by two distinct actors (as with the criminal act by the businessman and the pardon by the leader), but rather the action of the misbehaving leader in response to a decision by a different branch of government (the Court). This action is not described as criminal, but rather as a violation of political and institutional norms but in fact, this behavior is a crime in many jurisdictions.

We conducted a separate pre-test among a sample of 896 respondents using a convenience sample (Lucid) who were asked to rank order the seriousness of various behaviors that could be thought as deviations from ethical and democratic norms. All items were re-coded on a 0-1 scale with "1" reflecting the highest ranking in terms of crime seriousness. Investigating judges as a retaliation for their rulings had a mean score of 0.599 in terms of seriousness, higher than pardoning a party insider (Mean=0.536), which captures the violation in Studies 1 and 3. Only self-dealing (awarding contracts to his own firm) was a very close second (Mean=0.5881). Appendix Tables A12a and A12b provide details on all six items included in the survey and relevant statistics. This pre-test highlights that investigating judges is seen as more serious

[16] An alternative interpretation is that there is a threshold dynamic in effect: partisans' response to leaders may turn negative only for very serious crimes, more serious than those contemplated in these experiments. This interpretation, that partisans are fully complicit, is the more pessimistic scenario.

than pardoning party insiders, making the leader's behavior in Study 2 a more serious norms violation than in Studies 1 and 3.

Evaluation of the Misbehaving Leader Prior to Dissent

First, we look at the evaluations of the misbehaving leader once respondents were informed of the pardon of the insider (Studies 1 & 3) or the call for an investigation of judges (Study 2) but prior to any reaction from the co-partisan Assembly leader. For this analysis, we focus on two measures: affect for the misbehaving leader (thermometer score) and how they recommend that party members should vote in the primary.

Upon reading the story about the misbehaving leader's misbehavior, partisans expressed greater affect for the partisan Governor than the Australian Premier, which suggests an ingroup leadership bias consistent with social identity theory (Tajfel and Turner 1979).[17] Figure 2 shows the means for the misbehaving leader thermometer score in the Australia condition and the ingroup Governor condition for all three experiments (as well as evaluations of Trump in Study 3). In Study 1, affect for the misbehaving leader is higher in the ingroup condition compared to the Australia control ($Mean_G$=0.325; $Mean_A$=0.278; $p<0.001$). The pattern carries over in Study 2 which involves the more serious violation of interfering with the courts ($Mean_G$=0.404; $Mean_A$=0.374; $p=0.014$). In Study 3, conducted among Republicans, there is no ingroup favoritism effect for the Republican Governor relative to the Australian Premier, but there is a substantial difference of 31 percentage points when it comes to Trump who is rated far more warmly than either of the two hypothetical leaders ($Mean_G$=0.325; $Mean_A$=0.346; $Mean_T$=0.657; $p<0.001$). These patterns suggest

[17] Given the experimental design, we cannot say whether the extent of ingroup bias to a misbehaving leader is smaller, larger, or the same magnitude as the ingroup bias we would find if the leaders did not engage in norms violations. The differences in means presented here speak only to the presence of ingroup bias in the response to the misbehaving leader.

that partisans do not evaluate ingroup leaders identically to outgroup leaders, providing ingroup leaders with more leeway for norms violations. Moreover, in the case of the real ingroup leader, Trump, support among Republicans after exposure to the information about misbehavior far outpaces that for the hypothetical leaders.

[FIGURE 2- HERE]

Next, we assess the effect of the vignette on how respondents think party members should vote in the party primary – for the misbehaving leader or his challenger (Figure 3).[18] In Study 1, which represents the less serious violation case, we find that electoral support for the misbehaving leader is somewhat higher in Australia than in the partisan Governor condition ($Mean_G$=0.272; $Mean_A$=0.299; p=0.09), but this difference is not statistically significant at conventional levels (i.e., p<0.05). The difference is more pronounced in Study 2 which references the more serious normative violation (investigation of judges). Here, however, intent to vote for the neutral outgroup violator is significantly higher than for the ingroup partisan Governor ($Mean_G$=0.404; $Mean_A$=0.437; p=0.01). This suggests some optimism that ingroup leaders may be held accountable. However, Study 3 presents a starkly different picture. Although voting support for the misbehaving leader is similar between the Australia and the hypothetical Governor conditions ($Mean_G$=0.309; $Mean_A$=0.306; p=0.895), which is consistent with our findings in Study 1, this is not the case for Trump. Intent to vote for Trump is actually more than double of that for the control condition ($Mean_T$=0.653; $Mean_A$=0.306; p<0.001), providing little evidence that ingroup leaders are likely to be punished for norms violations.

[FIGURE 3-HERE]

[18] We opted to ask about voting in the *primary* rather than the *general* election because we expect that given the psychological pull of ingroup identity and partisanship in American elections, partisans are unlikely to switch their vote to a different party. It is more realistic to expect that if partisans are likely to punish a misbehaving leader, this would happen in the primary election.

From these results, we can draw several inferences. Affect and voting behavior do not always move in tandem; it depends both on who is being evaluated (hypothetical vs. real leader) and, to some extent, the seriousness of the infraction. In each of the studies, partisans exhibit ingroup bias in their affective evaluations (with Study 3 showing ingroup bias for President Trump only). Yet, we do not observe an ingroup bias in primary voting in any of the Governor conditions; only with Republicans and President Trump. The only evidence of partisans holding ingroup leaders more accountable than outgroup leaders is in the vote intention in the high seriousness condition (Study 2). Combined, this suggests that ingroup leaders may have leeway for moderate norms violations.[19]

Evaluation of the Misbehaving Leader after Ingroup Dissent

[19] Studies 1 and 3 included two items seeking to assess whether partisans believed that the misbehaving leader's response was justified because the pardon decision falls within his discretionary authority and because the businessman being pardon had proven his loyalty to the party (see Appendix A, Section A8 for the actual wording). We created an index based on these two binary items to assess to what degree there is an ingroup bias in offering a motivated justification for the leader's action. Results show that respondents assigned to the partisan Governor condition were significantly more likely to suggest that the Governor's actions were justified compared to the Australia control ($Mean_A$=0.179, $Mean_G$=0.229; $p<0.01$). In Study 3, we see no ingroup bias in motivated justification of the leader's action when it comes to the partisan Governor ($Mean_A$=0.213, $Mean_G$=0.236; p=0.493) but we do find a substantively large and statistically significant difference in willingness to excuse the action when it comes to Trump ($Mean_A$=0.213, $Mean_T$=0.421; $p<0.001$). In Study 2, we included an item suggesting that the misbehaving leader would be justified to shut down courts under certain circumstances (measured on a 4pt scale from very true to very false). We find no ingroup bias in motivated justification when it comes to the investigation of judges action ($Mean_A$=0.439, $Mean_G$=0.426; p=0.402) as a surprisingly large proportion of people in both groups agreed with the statement.

Next, we look at partisans' responses to the same two measures of affect and intent to vote, but after they were exposed to the reaction from the Assembly leader. These tests speak to whether dissent reduces support for the leader, as would be expected under our "instrumental" hypothesis, and the comparison between the Australia and ingroup Governor conditions allows us to examine if these responses are unique to ingroup leaders or translate more broadly to political leaders. Figure 4 shows the mean thermometer scores for Australia, the Governor, and Trump by loyalty and dissent from the Assembly Leader. In Study 1 (soft norms violation, pardon of insider), affect for the misbehaving Australian Premier is significantly stronger among those exposed to the dissent than the loyalty condition ($Mean_{Loyalty}$ = 0.285, $Mean_{Dissent}$ = 0.650, $p<0.001$). The same boost is evident in the case of the ingroup partisan Governor ($Mean_{Loyalty}$ = 0.350, $Mean_{Dissent}$ = 0.643, $p<0.001$). This suggests that exposure to dissent can produce an affective "rally around the leader" effect which emerges even when the leader who is being criticized belongs to a neutral outgroup. In essence, respondents appear to behave as ingroup members who rally on the side of the criticized leader. However, a different pattern emerges when we look at the hard norms violation case (Study 2; judges). Here, dissent has no effect on affect for the misbehaving leader in the Australia condition ($Mean_{Loyalty}$ = 0.367, $Mean_{Dissent}$ = 0.382, $p=0.413$). However, in the ingroup Governor condition, dissent produces a statistically significant affective *penalty* ($Mean_{Loyalty}$ = 0.429, $Mean_{Dissent}$ = 0.375, $p=0.005$). In Study 3, which is based on the same soft norms/less serious violation as Study 1, the results also suggest an affective "rally around the leader" effect not only for Australia ($Mean_{Loyalty}$ = 0.377, $Mean_{Dissent}$ = 0.709, $p<0.001$), and the hypothetical Governor ($Mean_{Loyalty}$ = 0.384, $Mean_{Dissent}$ = 0.716, $p<0.001$), but also for Trump ($Mean_{Loyalty}$ = 0.533, $Mean_{Dissent}$ = 0.622, $p=0.004$).

[FIGURE 4- HERE]

Next, we look at the effect of dissent on intent to vote for the misbehaving leader (Figure 5). In Study 1, representing the soft norms violation, the results suggest that dissent has no effect on partisans' intent to vote for the leader in the primary. This is the case for both the control condition of Australia ($Mean_{Loyalty}$ =

0.319, $Mean_{Dissent}$ = 0.323, p=0.845) and in the ingroup Governor condition ($Mean_{Loyalty}$ = 0.333, $Mean_{Dissent}$ = 0.337, p=0.870). Moving to Study 2, which references the hard norms violation, effect of dissent relative to loyalty is also null for both Australia ($Mean_{Loyalty}$ = 0.402, $Mean_{Dissent}$ = 0.417, p=0.423) and the partisan Governor condition ($Mean_{Loyalty}$ = 0.417, $Mean_{Dissent}$ = 0.386, p=0.102). The results are different in Study 3. Here, dissent leads to a significant "rally around the leader" effect for all three conditions, with respondents exposed to the dissent expressing much higher voting support for the Australian Premier ($Mean_{Loyalty}$=0.356; $Mean_{Dissent}$=0.740; p<0.001), the partisan Governor ($Mean_{Loyalty}$=0.405; $Mean_{Dissent}$=0.714; p<0.001), and Trump ($Mean_{Loyalty}$=0.472; $Mean_{Dissent}$=0.620; p<0.001). [20]

[FIGURE 5-HERE]

Taken together, these results suggest that dissent tends to either boost support for the misbehaving leader, or have no effect, whether affective or in terms of voting intent.[21] This suggests that people are not behaving as instrumental theories of democracy might suggest. Contrary to the ingroup leadership hypothesis, however, we found that even neutral outgroup leaders are not harmed by dissent.[22] Only in the case of the ingroup partisan leader involved in a very serious violation did we pick up an affective penalty,

[20] The results from Study 1 were partially replicated in a pilot study conducted on M-Turk which references the low seriousness condition only. Here too, we found that dissent can produce a "rally around the leader" effect, in this case, in the neutral control condition (Appendix B).

[21] We observe similar patterns using within-subjects analyses of affect and vote for the leader, controlling for initial assessments provided by respondents after the first portion of the vignette (see Appendix D).

[22] This result may be driven, to a degree, by question wording that induces our subjects to behave as ingroup members. The questions ask them to advise a member of the party in Australia on how to respond to their leader's misbehavior and the response from the co-partisan leader (see Appendix A, Section A8). It is possible that this may have activated some partisan ingroup considerations.

but since the experiment did not include a Trump condition, we are unable to say whether this extends to the case of real leaders with whom respondents have a stronger pre-existing affective attachment. Critically, not one of our experiments suggests an *electoral penalty* for the misbehaving leader when the Assembly leader dissents, regardless of the seriousness of the crime.

Evaluation of the Dissenting Co-Partisan Assembly Leader

Our results so far suggest that dissent has either no effect or the opposite effect from the one expected by instrumental perspectives and those that emphasize the public's commitment to democratic institutions. Could this be because partisans disapprove of dissent and perhaps think of it as evidence of inappropriate behavior? Our studies set out to assess this possibility by asking participants to evaluate the responding Assembly leader. We included three key measures: appropriateness of the response, feeling thermometer for the Assembly leader, and recommendations for how party members should in the Assembly leader's primary election. As Figure 6 shows, partisans generally assess dissent as an appropriate response to the Governor's misbehavior. This is true regardless of the seriousness of the infraction. Specifically, in Study 1, representing the soft norms violation, both in the Australia ($Mean_{Loyalty}$ =0.338, $Mean_{Dissent}$ =0.814, $p<0.001$) and the ingroup Governor conditions ($Mean_{Loyalty}$=0.415, $Mean_{Dissent}$=0.806, $p<0.001$) partisans judge the dissenting response as more appropriate than the loyalty response. This is also the case for Study 2, which deals with the more serious offense (Australia: $Mean_{Loyalty}$= 0.517, $Mean_{Dissent}$=0.732, $p<0.001$; Governor: $Mean_{Loyalty}$=0.570, $Mean_{Dissent}$ = 0.788, $p<0.001$). The same results emerge in Study 3, which includes the real leader as well as the hypothetical leaders (Australia: $Mean_{Loyalty}$=0.392; $Mean_{Dissent}$=0.821, $p<0.001$; Governor: $Mean_{Loyalty}$=0.460; $Mean_{Dissent}$=0.803, $p<0.001$; Trump: $Mean_{Loyalty}$=0.549; $Mean_{Dissent}$=0.715, $p<0.001$). Across the board, even in the case of the real partisan leader (in this case, Mitch

McConnell is the dissenter), dissent is judged as appropriate.[23]

[FIGURE 6-HERE]

Next, we look at feeling thermometer evaluations to assess whether affect for the Assembly leader is influenced by his dissent against the misbehaving leader. Results are shown in Figure 7. For Study 1 (soft norms), although the partisan dissenter's response is judged as more appropriate than the partisan loyalist's response, dissent seems to have no statistically discernible effect on affective evaluations, whether in the Australia condition ($Mean_{Loyalty}$=0.267, $Mean_{Dissent}$ = 0.274, p=0.705) or the ingroup Governor condition ($Mean_{Loyalty}$=0.316, $Mean_{Dissent}$ = 0.309, p=0.725). This is not the case, however, for Study 2 (hard norms). Here, where the gravity of the offense is greater, in both the Australia control ($Mean_{Loyalty}$ = 0.371, $Mean_{Dissent}$=0.569, p<0.001) and the ingroup Governor ($Mean_{Loyalty}$=0.449, $Mean_{Dissent}$=0.618, p<0.001) conditions, dissent boosts affect for the responding leader. In Study 3, we replicate the results from Study 1 which involves the same soft norms violation. In all three conditions, there is no difference in affect between the loyalist and the dissenter (Australia: $Mean_{Loyalty}$=0.340, $Mean_{Dissent}$ = 0.288, p=0.102; Governor: $Mean_{Loyalty}$=0.333, $Mean_{Dissent}$ = 0.321, p=0.729; Trump: $Mean_{Loyalty}$=0.661, $Mean_{Dissent}$ = 0.661, p=0.996). This suggests that only in the case of very serious offenses by the leader does co-partisan dissent boost affect for the critic.

[FIGURE 7-HERE]

Finally, Figure 8 shows the effects of dissent on primary vote for the co-partisan Assembly leader.

[23] Additional analyses of items asking whether the misbehaving leader, the Assembly leader, both or neither "best serves" the party and "embarrasses" the party also show that in the dissent conditions the Assembly leader is substantially more likely to be thought to "best serve" the party, while the Governor is more likely to "embarrass" the party (See Appendix A, Section A7). This further suggests that partisans see value in dissent and dislike leader misbehavior.

In Study 1, there is no difference between intent to vote for the loyalist and the dissenter either in the Australia control ($Mean_{Loyalty}$=0.512, $Mean_{Dissent}$ = 0.501, p=0.670) or in the ingroup Governor conditions ($Mean_{Loyalty}$=0.513, $Mean_{Dissent}$ = 0.500, p=0.547). When the norms violation by the leader is greater, however, we do observe dissent being rewarded. In Study 2 (hard norms; judges), dissent boosts primary voting for the responding Assembly leader in both the Australia control ($Mean_{Loyalty}$ = 0.426, $Mean_{Dissent}$=0.599, p<0.001) and the ingroup Governor conditions ($Mean_{Loyalty}$ = 0.476, $Mean_{Dissent}$ = 0.617, p<0.001). In Study 3, which also involves the soft norms violation, we see a replication of the null results from Study 1 for all three groups (Australia: $Mean_{Loyalty}$=0.472, $Mean_{Dissent}$ = 0.494, p=0.595; Governor: $Mean_{Loyalty}$=0.439, $Mean_{Dissent}$ = 0.498, p=0.125 Trump: $Mean_{Loyalty}$=0.520, $Mean_{Dissent}$ = 0.487, p=0.396). [24]

[FIGURE 8-HERE]

Taken together, our results suggest that partisans recognize dissent when they see it and deem it appropriate. In the case of serious leader violations, the dissenter also may get a boost in affect and even an electoral benefit. Yet, this same benefit does not hold in the case of less serious norms violations. When taken together with the findings that people rally around the leader when he is criticized by the Assembly leader in his own party, the results suggest that leaders may be immune to even ingroup criticism unless the misbehavior is deemed extremely serious. Moreover, criticism of the leader helps the dissenting co-partisan only in very extreme cases, giving them little reason to risk dissent against a popular leader.

Conclusion

Various perspectives suggest that dissent could be a powerful tool to enforce leader accountability, as voters are expected to update their evaluation of the leader based on the new information. Recent

[24] Appendix C provides power calculations that show that our studies had the power to detect relatively modest effects (e.g., one-fifth of a standard deviation), suggesting that these are fairly precisely estimated null results.

scholarship, taking into account the strong hold of partisanship on voter attitudes, has suggested that the most effective and believable source of dissent is from within the party. If fellow ingroup leaders identify a top leader's behavior as inappropriate or a violation of norms or laws, then voters will punish the misbehaving leader. Yet our studies show that even when dissent and criticism of the leader's actions originates from within the ingroup, this does not generally have the desired results. Even though people recognize dissent as appropriate, they generally fail to apply it in their evaluations of the misbehaving leader.

These patterns confirm that leaders occupy a special position within political groups, possibly as the key representations of group identity. Information about the leader that is inconsistent with the group's core values and norms may be threatening to group members because it endangers the normative boundaries and cohesion of the group. Under such circumstances, group members are motivated to discount the information and protect the leader. Yet we also find that people rally around leaders of neutral outgroups when faced with co-partisan dissent, which suggests that the power of leaders may extend beyond their ties to people's own group identities.

As shown in the Milgram (1963) studies, people have a strong tendency to acquiesce to authority even at the expense of others. This suggests that accountability may be limited in intragroup contexts or even in the case of neutral outgroup leaders. While serious norms violations and dissent from within the party can shift affective evaluations, a leader who violates established norms is unlikely to face an electoral penalty from the party base, even when the violation is considered to be quite serious. We do not know whether there is an upper bound to this tolerance of a norms-violating leader; however, our results show that dissent to both pardoning party insiders and politically motivated interference with the judicial system fail to significantly shift the electoral prospects of the leader in the primary election, even when they do shift affective attitudes.

Contemporary American politics further suggests that if leaders have unmediated access to their followers, for example through social media, not only can they defend themselves against legitimate ingroup

dissent, but perhaps force penalties on the dissenters. We have witnessed Donald Trump levying criticism against internal critics from Senator McCain, to James Comey, to Rep. Amash, to Senator Romney, to Administration officials testifying as part of the impeachment hearings. This subsequent step of leader response or backlash on dissenters is an open question that should be addressed in follow-up studies.

More broadly, this research suggests that the role of public opinion in the elevation and preservation of illiberal leaders through democratic electoral processes is a crucial area of research that has been sidelined by scholars of democracy and authoritarianism whose main focus has been institutions. The patterns found across the studies in this article suggest that political science needs to rethink mechanisms of leader accountability, especially in polarized democracies. Partisan social identity and extreme polarization may be threats to institutions in more ways than political scientists have anticipated so far.

Figure 1. Experimental Design (Studies 1, 2, 3)

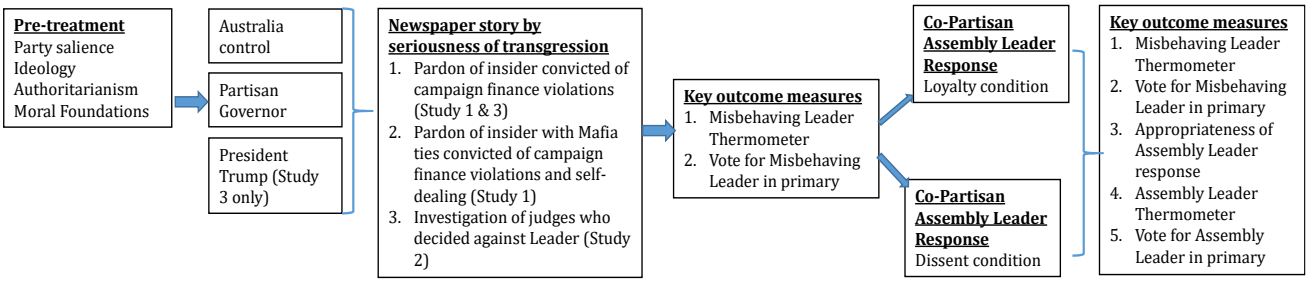

Figure 2. Affect for Governor (Prior to Assembly Leader Response)

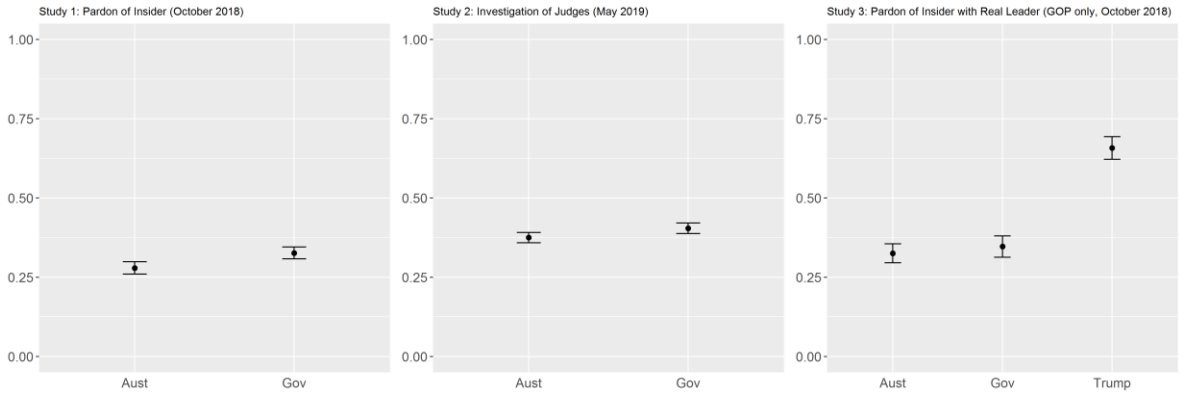

Figure 3. Intent to Vote for Governor (Prior to Assembly Leader Response)

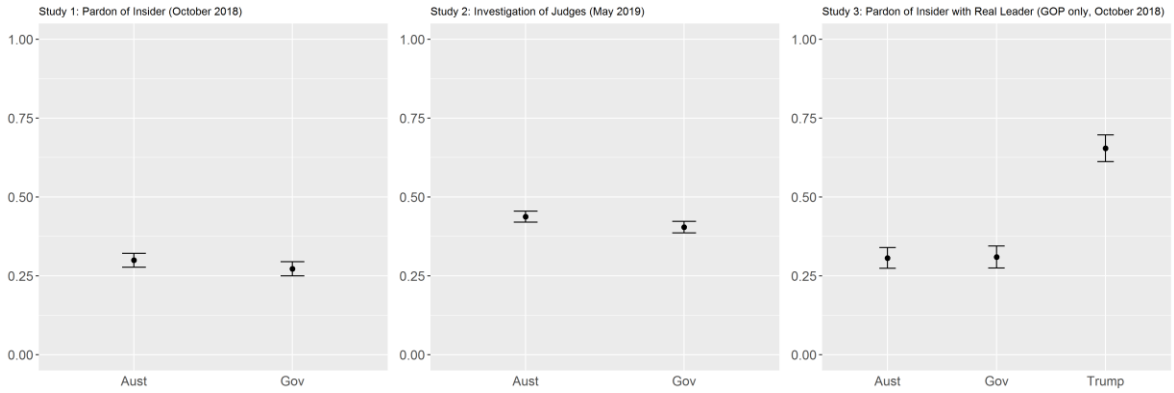

Figure 4. Affect for Governor (Post Assembly Leader Response)

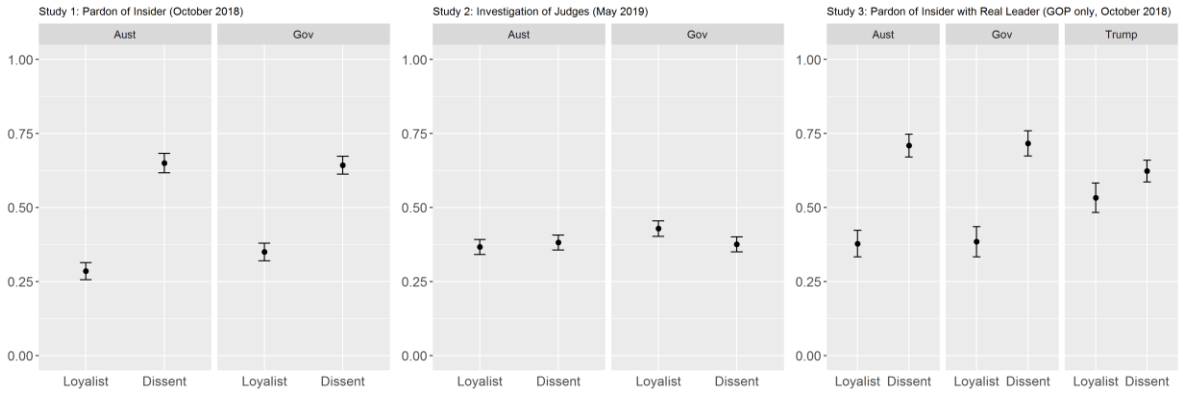

Figure 5. Intent to Vote for Governor (Post Assembly Leader Response)

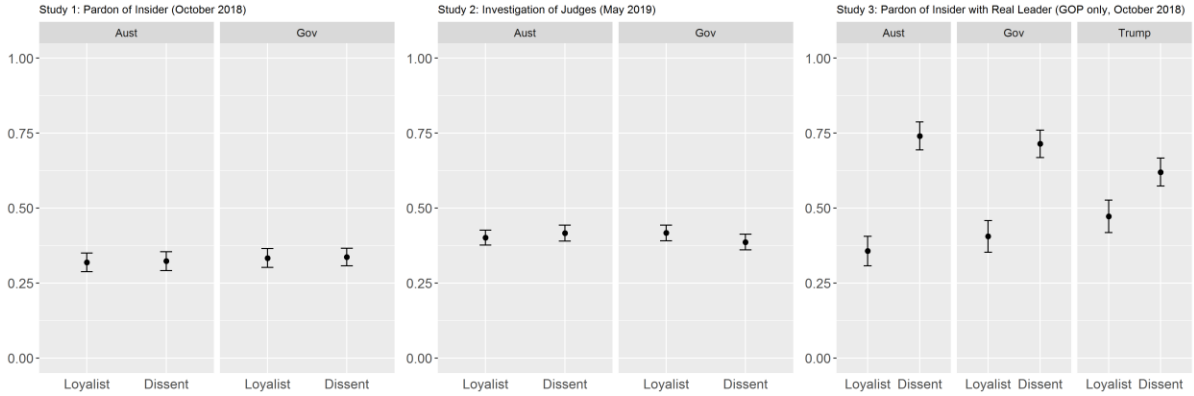

Figure 6. Appropriateness of Assembly Leader Response

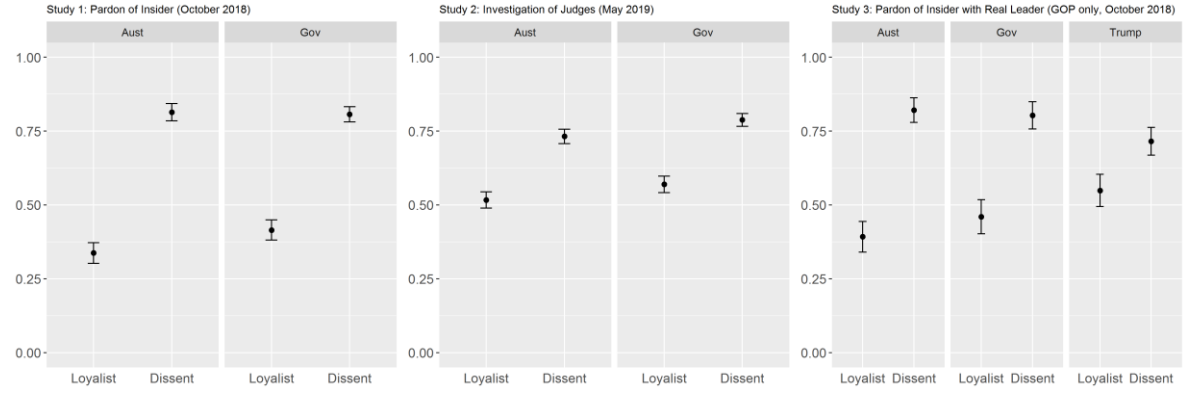

Figure 7. Affect for Responding Assembly Leader

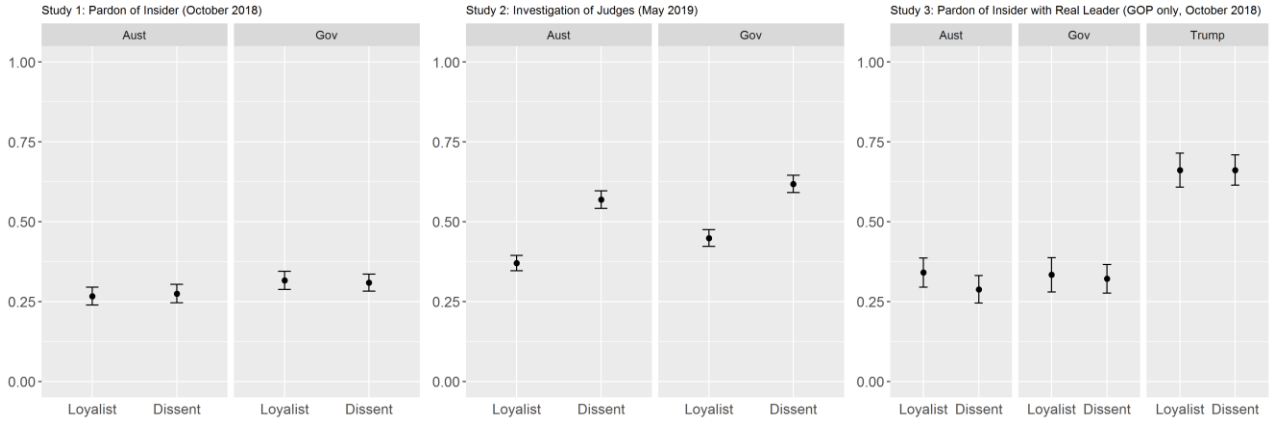

Figure 8. Intent to Vote for Assembly Leader

References

Abramowitz, Alan I., and Kyle L. Saunders. 2008. "Is Polarization a Myth?" *Journal of Politics* 70 (2):542-555.

Abrams, Dominic, Adam Rutland, Jennifer M. Ferrell, and Joseph Pelletier. 2008. "Children's Judgments of Disloyal and Immoral Peer Behavior: Subjective Group Dynamics in Minimal Intergroup Contexts." *Child Development* 79 (2):444-461. doi: doi:10.1111/j.1467-8624.2007.01135.x.

Barber, Michael, and Jeremy C. Pope. 2019. "Does Party Trump Ideology? Disentangling Party and Ideology in America." *American Political Science Review* 113 (1):38-54. doi: 10.1017/S0003055418000795.

Branscombe, Nyla R., Daniel L. Wann, Jeffrey G. Noel, and Jason Coleman. 1993. "In-Group or Out-Group Extemity: Importance of the Threatened Social Identity." *Personality and Social Psychology Bulletin* 19 (4):381-388. doi: 10.1177/0146167293194003.

Carey, John, Katherine Clayton, Gretchen Helmke, Brendan Nyhan, Mitchell Sanders, and Susan Stokes. 2019. "Who Will Defend Democracy? Evaluating Tradeoffs in Candidate Support Among Partisan Donors and Voters." *(unpublished manuscript)*.

Carey, John M., Gretchen Helmke, Brendan Nyhan, Mitchell Sanders, and Susan Stokes. 2019. "Searching for Bright Lines in the Trump Presidency." *Perspectives on Politics* 17 (3):699-718.

Chekroun, Peggy, and Armelle Nugier. 2011. ""I'm ashamed because of you, so please, don't do that!": Reactions to deviance as a protection against a threat to social image." *European Journal of Social Psychology* 41 (4):479-488. doi: doi:10.1002/ejsp.809.

Delli Carpini, Michael, and Scott Keeter. 1996. *What Americans Know About Politics and Why It Matters*. New Havern, CT: Yale University Press.

Fiorina, Morris P. 1981. *Retrospective Voting in American National Elections*. New Haven, CT: Yale University Press.

Foa, Roberto Stefan, and Yascha Mounk. 2016. "The Democratic Disconnect." *Journal of Democracy* 27 (3):5-17.

Galinsky, A. D., J. C. Magee, D. H. Gruenfeld, J. A. Whitson, and K. A. Liljenquist. 2008. "Power reduces the press of the situation: Implications for creativity, conformity, and dissonance." *Journal of Personality and Social Psychology* 95 (6):1450-1466.

Giessner, Steffen R., Daan van Knippenberg, and Ed Sleebos. 2009. "License to fail? How leader group prototypicality moderates the effects of leader performance on perceptions of leadership effectiveness." *The Leadership Quarterly* 20 (3):434-451. doi: https://doi.org/10.1016/j.leaqua.2009.03.012.

Graham, Matthew, and Milan Svolik. 2019. "Democracy in America? Partisanship, Polarization and the Robustness of Support for Democracy in the United States." *(unpublished manuscript)*.

Haslam, S. Alexander, Michael J. Platow, John C. Turner, Katherine J. Reynolds, Craig McGarty, Penelope J. Oakes, Susan Johnson, Michelle K. Ryan, and Kristine Veenstra. 2001. "Social Identity and the Romance of Leadership: The Importance of being Seen to be 'Doing it for Us'." *Group Processes & Intergroup Relations* 4 (3):191-205. doi: 10.1177/1368430201004003002.

Hetherington, Marc. 2005. *Why Trust Matters*. Princeton, NJ: Princeton University Press.

Hogg, M.A. 1992. *The Social Psychology of Group Cohisiveness: From Attraction to Social Identity*. London: Harvester Wheatsheaf.

Hogg, Michael A. 2001. "A Social Identity Theory of Leadership." *Personality and Social Psychology Review* 5 (3):184-200. doi: doi:10.1207/S15327957PSPR0503_1.

Hornsey, Matthew J., and Jolanda Jetten. 2003. "Not being what you claim to be: impostors as sources of group threat." *European Journal of Social Psychology* 33 (5):639-657. doi: doi:10.1002/ejsp.176.

Hornsey, Matthew J., Jolanda Jetten, Brendan J. McAuliffe, and Michael A. Hogg. 2006. "The impact of individualist and collectivist group norms on evaluations of dissenting group members." *Journal of Experimental Social Psychology* 42 (1):57-68. doi: https://doi.org/10.1016/j.jesp.2005.01.006.

Hornsey, Matthew J., Mark Trembath, and Sasha Gunthorpe. 2004. "'You can criticize because you care': identity attachment, constructiveness, and the intergroup sensitivity effect." *European Journal of Social Psychology* 34 (5):499-518. doi: 10.1002/ejsp.212.

Huddy, Leonie, Lilliana Mason, and Lene Aaroe. 2015. "Expressive Partisanship: Campaign Involvement, Political Emotion, and Partisan Identity." *American Political Science Review* 109 (1):1-17.

Hutchison, Paul, Dominic Abrams, Roberto Gutierrez, and G. Tendayi Viki. 2008. "Getting rid of the bad ones: The relationship between group identification, deviant derogation, and identity maintenance." *Journal of Experimental Social Psychology* 44 (3):874-881. doi: https://doi.org/10.1016/j.jesp.2007.09.001.

Iyengar, Shanto, Gaurav Sood, and Yphtach Lelkes. 2012. "Affect, Not Ideology: A Social Identity Perspective on Polarization." *Public Opinion Quarterly* 76 (3):405-431. doi: 10.1093/poq/nfs038.

Iyengar, Shanto, and Sean J. Westwood. 2015. "Fear and Loathing across Party Lines: New Evidence on Group Polarization." *American Journal of Political Science* 59 (3):690-707. doi: 10.1111/ajps.12152.

Jennings, Kent M., and Richard G. Niemi. 1974. *The Political Character of Adolescence*. Princeton, NJ: Princeton University Press.

Jetten, Jolanda, and Matthew Horsney. 2014. "Deviance and Dissent in Groups." *Annual Review of Psychology* 65:461-85.

Katz, Josh, and Kevin Quealy. 2017. "Which Country is America's Strongest Aly? For Republicans, It's Australia." *The New York Times* (February 3, 2017) https://www.nytimes.com/interactive/2017/02/03/upshot/which-country-do-americans-like-most-for-republicans-its-australia.html.

Levitsky, Steven, and Daniel Ziblatt. 2017. *How Democracies Dies*. New York: Crown Publishing.

Marques, José M., Vincent Y. Yzerbyt, and Jacques-Philippe Leyens. 1988. "The “Black Sheep Effect”: Extremity of judgments towards ingroup members as a function of group identification." *European Journal of Social Psychology* 18 (1):1-16. doi: 10.1002/ejsp.2420180102.

Mason, Lilliana. 2018. *Uncivil Agreement: How Politics Became Our Identity*. Chicago, IL: University of Chicago Press.

Milgram, Stanley. 1963. "Behavioral Study of obedience." *The Journal of Abnormal and Social Psychology* 67 (4):371-378. doi: 10.1037/h0040525.

Mishler, William, and Richard Rose. 2001. "What Are the Origins of Political Trust?:Testing Institutional and Cultural Theories in Post-communist Societies." *Comparative Political Studies* 34 (1):30-62. doi: 10.1177/0010414001034001002.

Monin, Benoît, Pamela J. Sawyer, and Matthew J. Marquez. 2008. "The rejection of moral rebels: Resenting those who do the right thing." *Journal of Personality and Social Psychology* 95 (1):76-93. doi: 10.1037/0022-3514.95.1.76.

Nicholson, Stephen P. 2012. "Polarizing Cues." *American Journal of Political Science* 56 (1):52-66. doi: doi:10.1111/j.1540-5907.2011.00541.x.

Rast, David E., Justin D. Hackett, Alexis Alabastro, and Michael A. Hogg. 2015. "Revoking a leader's “license to fail”: downgrading evaluations of prototypical in-group leaders following an intergroup failure." *Journal of Applied Social Psychology* 45 (6):311-318. doi: doi:10.1111/jasp.12297.

Redlawsk, David P. 2002. "Hot Cognition or Cool Consideration? Testing the Effects of Motivated Reasoning on Political Decision Making." *The Journal of Politics* 64 (4):1021-1044. doi: 10.2307/1520074.

Reicher, Stephen, and S. Alexander Haslam. 2011. "After shock? Towards a social identity explanation of the Milgram 'obedience' studies." *British Journal of Social Psychology* 50 (1):163-169. doi: 10.1111/j.2044-8309.2010.02015.x.

Reicher, Stephen, S. Alexander Haslam, and Rakshi Rath. 2008. "Making a Virtue of Evil: A Five-Step Social Identity Model of the Development of Collective Hate." *Social and Personality Psychology Compass* 2 (3):1313-1344. doi: 10.1111/j.1751-9004.2008.00113.x.

Rosenblum, Nancy L. 2008. *On the Side of the Angels*. Princeton, NJ: Princeton University Press.

Rothstein, Bo, and Auiysha Varraich. 2017. *Making Sense of Corruption*. New York: Cambridge University Press.

Steffens, Niklas K., S. Alexander Haslam, and Stephen D. Reicher. 2014. "Up close and personal: Evidence that shared social identity is a basis for the 'special' relationship that binds followers to leaders." *The Leadership Quarterly* 25 (2):296-313. doi: https://doi.org/10.1016/j.leaqua.2013.08.008.

Steffens, Niklas K., Sebastian C. Schuh, S. Alexander Haslam, Antonia Pérez, and Rolf Dick. 2015. "'Of the group' and 'for the group': How followership is shaped by leaders' prototypicality and group identification." *European Journal of Social Psychology* 45 (2):180-190. doi: 10.1002/ejsp.2088.

Svolik, MIlan W. 2019. "Polarization versus Democracy." *Journal of Democracy* 30 (3):20-32.

Taber, Charles S., and Milton Lodge. 2006. "Motivated Skepticism in the Evaluation of Political Beliefs." *American Journal of Political Science* 50 (3):655-769.

Tajfel, Henri, and Jonathan Turner. 1979. "An Integrative Theory of Intergroup Conflict." In *The social psychology of intergroup relations*, 33-47. Taylor and Francis Psychology Press.

Walter, Annemarie S., and David P. Redlawsk. 2019. "Voters' Partisan Responses to Politicians' Immoral Behavior." *Political Psychology* 0 (0). doi: 10.1111/pops.12582.

Wood, Abby K., and Christian R. Grose. 2019. "Campaign Finance Transparency Affects Legislators' Election Outcomes and Behavior." *(working paper)*.

Zaller, John. 1992. *The Nature and Origins of Mass Opinion*. Cambridge, UK: Cambridge University Press.