Who is it Okay to Punch? An Experimental Investigation of Support for Intolerance in the Form of Physical Violence

Frank J. Gonzalez
Assistant Professor
School of Government and Public Policy
University of Arizona
fgonzo@arizona.edu

Alexandra McCoy
MA Student
Arizona State University
ajmccoy1@asu.edu

A popular contemporary meme involves the social acceptability of “punching Nazis.” This phenomenon raises the question: what characteristics make a group or member of a group more or less “punch-able”? More broadly, what group attributes yield support for physical violence against them? In this study, we build on the extant psychology literature on intergroup processes to look at what leads individuals to find physical violence against a group acceptable, and if the factors that lead to such acceptance differ from those that lead to sheer affective intolerance. We use two experimental tasks to test expectations built on prominent theories. In a “real-world rating task,” participants evaluated a series of real-world groups and individuals with varying characteristics such as race, ideology, intelligence, warmth, and tendency toward violence. Also, in a conjoint experiment, participants chose between two groups in terms of which they would support being punched, as various attributes were manipulated such as race, expressed ideological values, partisanship, income, tendency toward violence, and being described as disgusting or threatening. We find consistent support for effects of political alignment (particularly, expressed ideological values related to dominance) and perceived tendency toward violence, as well as mixed findings for race and several other factors. These findings synthesize theories on intergroup conflict with contemporary findings on affective polarization.

Keywords: Political intolerance, intergroup processes, political violence, affective polarization, conjoint analysis

In January of 2017, white nationalist and self-proclaimed founder of the Alt-Right, Richard Spencer, was punched in the face while chanting, “Hail Trump! Hail our people! Hail victory!” against protestors of former president Donald Trump’s inauguration (Murphy, 2017). This “sucker-punch” of a controversial figure sparked a national discussion on the acceptability of political violence, often centered on social media around memes related to “punching Nazis” (Lopez, 2017). Many argued this type of violence should be vehemently opposed in America, as political issues should be resolved through “civil discussions, peaceful protests, and the vote” and that hurt one’s own cause and reinforce the other side (Lopez, 2017). However, others argued there are times when violence is acceptable, if not necessary, as a means of suppressing groups and ideologies that promote illiberal ideas that would, ironically, suppress the freedom of expression and democratic values that allowed them to thrive in the first place (Lopez, 2017). To the victims of oppressive ideologies in particular, talk of civility and “respectability politics” is an unfair burden (Young, 2016).

Yet justifications of violence can go both ways, as was evident during the capitol riots of January 6th, 2021. Former president Donald Trump proclaimed to his supporters that, “We fight like hell and if you don’t fight like hell, you’re not going to have a country anymore” (Woodward, 2021). That very same day, a violent mob of individuals who were loyal to Donald Trump stormed the U.S. capital to overturn America’s presidential election, as well as to brutalize national representatives. While the January 6th attack on the capital may be the most extreme example, as it directly attacked democratic institutions, research suggests the adversarial nature of the two-party system in America in conjunction with identity-related psychological motivations has yielded a political climate ripe for tension and, at times, violence. However, there does seem to be some variation in the perceived acceptability of who it is okay to be violent

against. On the left, there seems to be a “civility pass” when it comes to outright Neo-Nazis, but calls for sucker-punching other right-wing figures, even those with more extreme views, seems rare. On the right, there is a greater tolerance for violence against some groups – e.g., foreign enemies, terrorists, criminals – and although the “sucker punching” conversation seems to not have manifested among conservatives, there has been infighting with regard to the appropriateness of violence against political figures (e.g., AOC, Nancy Pelosi), such as the attempted violence of the January 6th Capitol riots.¹ So, what makes some groups perceived as more “punch-able” than others? What are the psychological mechanisms by which people evaluate when violence is okay against certain groups? Which perceived characteristics of a potential target matter most?

The goal of this study is to assess the explanatory power of factors we should expect to be relevant given existing theories of intergroup conflict and violence, as well as more recent work on affective polarization. Our hope is that a broad analysis of various potential factors will help to hone future examinations into what those critical factors are that lead some political groups to be identified as outside the boundaries of civility.

Affective Polarization and its Limits in American Politics

The literature of affective polarization suggests that there is a large division that has emerged in the mass public in recent years: “ordinary American citizens increasingly dislike and distrust those from the other party” (Iyengar et al., 2019). Citizens respond to ideological divergence with heightened affective polarization and these hostile feelings for the opposing party seem to be ingrained and automatic in voters’ minds (Iyengar et al., 2019; Iyengar &

¹Additionally, more extreme figures on the right-leaning fringe clearly support violence against the supposed baby-eating, Satanic cabal thought to exist in government.

Westwood, 2015; Rogowski & Sutherland, 2016). Affective polarization thought to be powerful enough to exert effects on non-political judgements and behaviors (Iyengar & Westwood, 2015), encourage partisan motivated reasoning (Martherus, Martinez, Piff, & Theodoridis 2019), influence responses to the Covid-19 pandemic (Druckman et al., 2021), and yield especially concerning phenomena such as dehumanization (Cassese, 2019) and support for violence (Kalmoe & Mason, 2019). If these latter manifestations of polarization grow, this poses a particular threat to democracy as reciprocal relationships can develop between intractable intergroup conflict and ideological polarization (Harel et al., 2020).

It is important to note that partisan prejudice is not dichotomous, but rather, exists on a continuum (see Alport, 1954). Some have argued that although partisan prejudice may cause incidental or direct harm to unfavorable outgroups, those with the highest rates of affective polarization are still no more willing to discriminate against or intentionally harm the opposition compared to those on the lower end of the spectrum (Lelkes & Westwood, 2017; Westwood & Lelkes, 2019). Even further, some work suggests the degree to which affective polarization exists in the American public in the first place is overstated (Druckman & Levendusky, 2019; Klar, Krupnikov, & Ryan, 2018). So, although saying partisans dislike each other might be an understatement, there does not seem to be much evidence that actual violence is widely supported. Yet we know that it happens, and although it does not seem to be violence based on the partisanship of a group alone, it does seem to be political. What might we learn from the psychology literature on intergroup conflict that can help us predict when political violence *is* likely to occur?

Political Intolerance – Psychological Mechanisms

Intergroup conflict is an incredibly broad topic in social psychology, but we focus on a few theories we think might be especially helpful in terms of predicting support for violence against political groups. (but see, e.g., Marcus et al., 1995 for a thorough application of intergroup conflict theories to political tolerance). The Stereotype Content Model (SCM) is well-known and provides a simple but robust framework of the influences that underlie stereotyping (Cuddy, Fiske, Glick, et. al., 2002). The SCM provides support that tolerance generally should be influenced by the perceived “competence” and “warmth” of the outgroup (Cuddy, Fiske, Glick, et. al., 2002; Fiske, 2012). For example, groups high in warmth and high in competence tend to elicit pride and admiration whereas groups who are low in warmth and low in competence tend to be associated with feelings of anger, resentment, and disgust (Cuddy, Fiske, Glick, et. al., 2002). Similarly, the Behaviors from Intergroup Affect and Stereotypes (BIAS) Map piggybacks off of the SCM in order to predict how the underlying fundamental dimensions of stereotypes – “warmth” and “competence” – lead to distinct emotions, which then lead to specific behaviors (e.g., Cuddy, Fiske, & Glick, 2007).

Some work based on the SCM and BIAS Map suggests the effects of competence should depend on the type of intolerance. Examining the evolution of stereotypes about Jews preceding and into the Holocaust, Bilewicz and Vollhardt (2013) examine the transformation of stereotypes about Jews from cold-competence (powerful threats) in the time leading up to the Holocaust to cold-incompetence (weak and disgusting vermin) during it. Cold-competent stereotypes are key to scapegoating, and these stereotypes can be leveraged by majority groups to “to legitimize their power and mobilize the ingroup against the stereotyped outgroup” (Bilewicz & Vollhardt 2013, p. 285; see also, Glick 2005). Cold-incompetent stereotypes elicit disgust and contempt, which

triggers avoidance. As such, if we are looking to predict which groups are most likely to be targets of violence and aggression, we should look to those perceived as cold yet competent.

Other intergroup conflict theories besides the BIAS Map predict different effects of perceived competence or formidability of the outgroup. Intergroup Emotions Theory (IET) suggests that appraisals of a given situation give rise to relative comparisons of group formidability, which then lead to distinct emotions and specific action tendencies (e.g., Mackie, Devos, & Smith, 2000). According to IET, after an ingroup feels threatened by an outgroup, the ingroup will assess their strength compared to the outgroup. If the outgroup is perceived as stronger, individuals will experience fear and avoid the outgroup, but aggress against weaker outgroups (Mackie, Devos, & Smith, 2000). Are political groups perceived as incompetent the most likely to be the targets of violence?

The Socio-Functional Threat model of prejudice takes a somewhat more nuanced approach and, of particular interest for this study, highlights the distinction between fear and disgust (Cottrell & Neuberg, 2005). According to this model, physical or moral contaminants elicit disgust, resulting in rejection behavior (e.g., Rozin, Haidt, & McCauley, 1999), whereas fear emerges when people are physically endangered, motivating escape behaviors (Ledoux, 1996). It may be the case, then, that groups characterized as disgusting are most likely to be the groups people supporting getting rid of, even using violent means. Groups that instill fear, on the other hand, should just provoke avoidance.

Factors to Investigate and Expectations for Each

Ultimately, we still don't have much in terms of comprehensive understandings of the characteristics of groups most likely to incur violence. However, we build on a combination of the intergroup conflict theories described above as well as several other literatures related to

intolerance (political or otherwise) to identify several factors we think are particularly ripe for empirical investigation.

Competence/Formidability. All the intergroup conflict theories described above mention a role for perceived competence or formidability in some way but come to somewhat different conclusions about how it might be related to violence. According to the SCM, those who are perceived as formidable as well as competent are usually treated in high regard. However, this positive association can fade if the perceived ‘warmth’ component is not present (Cuddy, Fiske, Glick, et. al., 2002). For example, a 2017 meta-analysis looking at racial bias in judgment of physical size and formidability found that people unjustly perceive young black adolescents as bigger (heavier, taller, muscular) as well as more physically threatening (overpowering, higher capability of harm) than white men in the same age demographic (Wilson, Hugenberg, & Rule, 2017). These biased formidability judgements can offer justifications for use of force against Black suspects of crime (Wilson, Hugenberg, & Rule, 2017). In more direct relation to the current study, Kalmoe and Mason (2019) show that the expectation of electoral victory gives strong partisans more confidence to endorse violence against political opponents. As such, a hypothesis derived from the SCM and BIAS Map may be:

SCM Hypothesis: Perceived competence/formidability is positively associated with support for violence against groups seen as “cold,” but negatively associated with support for violence against groups seen as “warm.”

On the other hand, the alternative intergroup conflict theories described suggest a potentially more straightforward hypothesis whereby the fear instilled by perceiving a group as formidable triggers avoid motivations, and aggression should be expected mainly when the target

is an outgroup and is seen as “defeat-able” (Cottrell & Neuberg, 2005; Mackie, Devos, & Smith, 2000). As such, we put forward a hypothesis to compete with the *SCM Hypothesis*.

Pray on the Weak Hypothesis: Perceived competence/formidability will be negatively associated with support for violence.

Disgust versus Threat. Disgust has been shown to be robustly related to politics in a variety of ways (e.g. Aarøe, Petersen, & Arceneaux, 2017; Inbar, Pizzaro, Iyer, & Haidt, 2012; Kam & Estes, 2016). In the SCM, disgust is elicited by cold-incompetence (Cuddy, Fiske, Glick, et. al., 2002). These effects are strong and automatic enough that some work suggests that the sight of homeless people (an example of a group seen as cold-incompetent) activates the insula in people’s brains, which is an area “implicated in disgust and arousal” (Harris & Fiske, 2006). These strong and innate neural patterns fit the “dehumanization of extreme outgroups in the SCM space” (Cuddy, Fisk, & Glick, 2006). Further, moral transgressions can elicit disgust, which suggests that moral cognition may draw upon a more primitive rejection response (Chapman & Anderson, 2013). This is contrasted with the threat elicited by cold-competent groups, which elicits aggression.

Disgust Avoidance Hypothesis: Perceiving a group as disgusting will be negatively associated with support for violence.

Yet again, other intergroup conflict literature suggests the contrary. That is, the fear provoked by powerful, threatening groups may trigger avoidance, whereas disgust may lead to a desire for expulsion of the potential contagion (Cottrell & Neuberg, 2005; Rozin, Haidt, & McCauley, 1999), and potentially, eradication.

Disgust Eradication Hypothesis: Perceiving a group as disgusting will be positively associated with support for violence.

Race. Race is another key factor that we should expect to be relevant here. In fact, when looking for manifestations of political violence, violence based on race is perhaps the most evident and egregious in American politics (e.g., Horowitz 1985; Posner 2004; Varshney 2003; Wilkinson 2004). Racial and ethnic minorities are historically the targets of prejudice, violence, scapegoating, and, in extreme cases, genocide (Tindale et al., 2002). As such, we might expect the following:

Straightforward Prejudice Hypothesis: Support for violence will be higher for groups comprised of racial and ethnic minorities compared to groups that are primarily White.

Yet an expansive literature in political psychology suggests this may not be the case. Due to societal norms, it has become less socially acceptable to be overtly racist and overt racial cues (rather than subtle, covert cues) may trigger motivations to avoid appearing prejudiced (e.g., Mendelberg, 2001; Valentino, Hutchings, and White, 2002; but see Valentino, Neuner, & Vandebroek, 2018). As a result of this, groups mentioned as being comprised of racial minorities may be somewhat protected by this norm. Although, we acknowledge that in general, we *know* that minorities are more likely to be targets of violence. However, egalitarian norms may provide a buffer against expressed support for violence.

Egalitarian Norms Hypothesis: Support for violence will be higher for groups that are primarily White compared to groups comprised of racial minorities.

Support for Violence. A somewhat surprisingly overlooked factor in existing literature is perceptions that the target group is, themselves, violent. In some sense, this makes straightforward sense in line with the “norm of reciprocity” (albeit the dark side of this norm). “An eye for an eye” logic has been shown to underlie political violence in a comparative sense (e.g., Fielding & Shortland, 2010; Liberman, 2006) as well as in organizational psychology (e.g., Greco et al., 2019). Perceived violence of an outgroup is seen as a justification for foregoing deliberation and using force. There are important intellectual, existential, and physical needs that tend to motivate individuals to become ideologically extreme (Webber et al., 2020). To serve these needs, there are various mechanisms that elicit and promote political violence, including ideologies that can “moralize values and portray violence as a remedy to collective grievances” (Webber et al., 2020). Additionally, support for intergroup violence can be explained in terms of “whether or not one accepts the political ends towards which violence is being used” (Henry et al., 2005; Kruglanski et al., 2014) as well as how violence is normalized in these contexts (See Hayes & McAllister, 2001; Harff, 2003; Tindale, Munier, Wasserman, & Smith, 2002).

Violence Hypothesis: Support for violence will be higher for groups perceived as violent compared to groups perceived as nonviolent.

Political Alignment. It is clear that partisanship is a strong social identity (See Huddy, Mason, and Aaroe 2015; Iyengar et al. 2012; Iyengar and Westwood 2015). Affective polarization seems to be the strongest among individuals with stronger ideological commitments and higher levels of political interest (Rogowski & Sutherland, 2016), and the emotions that result from group conflict are likely to be heightened among well-sorted partisans (Mason, 2016). Further, political

intolerance may in fact be more ‘natural’ than political tolerance (McClosky & Brill, 1983), as it is easier for individuals to be intolerant in the political domain because of the ease with which people acquire stereotypes and prejudices towards those not in their ingroup (e.g., Abouand, 1988; Devine, 1989). Indeed, partisan and ideological stereotypes are pervasive in American politics (Clifford, 2020; Crawford, Modri, & Motyl, 2013). Additionally, ideological values seem to conflict often in American politics (e.g., Lupton, Smallpage, & Enders, 2020). The Ideological-Conflict Hypothesis states this most clearly, arguing that partisans are partial towards their worldview, and when there are threats to said worldview, they will engage in various strategies to reject these threats (e.g., Brandt et al., 2014). Unsurprisingly, individuals are less likely to support freedom of expression if they dislike the outgroup, especially if they consider the outgroup as a moral violation or as a threat to one’s own ingroup (Marcus et al., 1995). As such, it may be the case that groups perceived as representing the “other side” politically are most likely to be the targets of violence.

Political Alignment Hypothesis: Perceived political distance (in terms of partisanship and ideology) will be positively associated with support for violence.

Importantly, we also set up two competing hypotheses within the framework of the general expectations of the *Political Alignment Hypothesis*. What does it tell us if political distance matters? Does this mean that sheer political labels are enough to provoke violence because of simple ingroup-outgroup dynamics? Or, do political labels matter because they reflect some more deep-seated conflict in values, morals, or worldview? In other words, an interesting question beyond whether political alignment matters is whether partisan labels or professed ideologies and values of the group matter more. We put forward the following hypotheses in this spirit.

Partisan Identity Hypothesis: Support for violence will be higher for groups labeled as the political outgroup than for groups labeled as the political ingroup.

Ideological Values Hypothesis: Support for violence will be higher for groups that express values inconsistent with one's own compared with groups that express values consistent with one's own.

Study Overview

Participants completed two experimental tasks: 1) a real-world target rating (RWTR) task, which required participants to rate a series of 16 real-world groups and individuals in terms of how positively/negatively they felt toward the group/individual (target), how upset they would be if the target was physically harmed, and several other dimensions, and 2) a conjoint rating task, which required participants to complete 30 trials in which they chose which of two groups that varied along several experimentally manipulated dimensions they would be “more likely to support violence against in the form of a member of that group being punched in the face while in public.” Although the RWTR task allows us to distinguish somewhat the predictors of feelings toward targets relative to attitudes toward the target being harmed, it is more limited than the conjoint rating task in its ability to distinguish some of the aforementioned hypotheses, as will be seen. We discuss the design and results of the RWTR task first, but the order of these tasks was randomized across participants.

Real-World Target Rating Task

Data

One hundred and seventeen undergraduate students enrolled in political science and public policy courses at a mid-sized southwestern university were recruited during academic

semesters between November 18, 2020 and March 9, 2021, to participate in the study in exchange for course credit. Twenty-two participants did not consent to have their data analyzed (consent was asked twice – once at the beginning and once at the end of the survey), and 5 responded to a self-report attention check question that they did not pay attention, leaving a final sample for analysis of 90 (56.7% White, 30% Latinx, 30% male, 90% English-as-first-language, median age = 20 (range = 18-49), median household income between \$80,000 and \$100,000, 71.1% Democratic, 16.7% Republican).²

Protocol and Experimental Design

Participants rated 8 real-world individuals and 8 real-world groups, which were chosen based on pre-testing. Whether individuals rated individuals or groups first was randomized across participants, as was the order of individuals/groups within each section. In a pre-test, a separate sample of online respondents rated 102 groups and individuals along several dimensions, including how liberal or conservative, white, violent, popular, intelligent, warm, and large (for groups only) they were. Based on these pre-test data, we chose 4 groups and 4 individuals to represent each of the following 8 categories: liberal, conservative, violent, nonviolent, intelligent, unintelligent, warm, and cold. Groups/individuals were chosen for the categories such that they were rated highly on that dimension but also such that the 4 groups/individuals chosen displayed variation on all *other* rated dimensions as well. Participants in the main study were randomly assigned one group and one individual representing each category to ensure variation in the targets being rated in terms of these dimensions.

² *F*-tests and chi-squared proportions tests suggested omissions were not significantly related to any of our experimental conditions.

Participants were first asked if they were familiar with the target (if they were not, they were not asked to rate that target), and then rated the target on 5-point scales along the following dimensions: how positive or negative they feel toward the target ($M = 2.80$, $SD = 1.40$), how upset they would be if the target were physically harmed ($M = 2.89$, $SD = 1.44$), and to what degree they perceived the target as warm (versus cold) ($M = 2.68$, $SD = 1.31$), intelligent (versus unintelligent) ($M = 3.46$, $SD = 1.18$), conservative (versus liberal) ($M = 3.33$, $SD = 1.21$), and violent (versus nonviolent) ($M = 3.16$, $SD = 1.38$). These 4 perceptions items were asked in randomized order.³ Race of the target was also coded by using the mean score for how ‘white’ the target was rated in the pre-test ($M = 3.71$, $SD = 1.14$, range = 1 to 5).⁴ Given these measures, we have the ability to reasonably assess support for the *Racial Prejudice* versus *Egalitarian Norms* hypotheses, the hypotheses related to perceived competence and formidability, the *Political Alignment Hypothesis*, and the *Violence Hypothesis*. This analysis also allows us to identify differences in the processes underlying feelings toward a target versus those underlying attitudes toward a target being physically harmed. We leave tests of the additional hypotheses to the conjoint analysis. See the Appendix for a description of the pre-test. All 64 targets and rating statistics available online through the Open Science Framework (<https://osf.io/vga2r/>).

³ We also asked attention check questions for each section (groups and individuals) where we asked participants to fill out specific responses (e.g., “‘No’ to whether or not you are familiar”, “‘Somewhat positive’ to how negative or positive you feel”) to a question where no individual or group was shown. Out of the 6 total attention check items, 73 participants got 0 wrong and none got more than 3 wrong. Given the relative complexity of these attention check items, we saw this as evidence that all participants paid reasonable levels of attention to the task and did not omit anyone from analyses.

⁴ It would also be possible to hand-code each target in a categorical manner based on their known race. However, given that the race of many targets (groups, in particular) is ambiguous, as well as the fact that perceptions of one’s whiteness or blackness are often more continuous than they are discrete (e.g., Ho, Sidanius, Levin, & Banaji, 2011; Khanna, 2010), we opted to use this continuous variable, which simply codes the degree to which the target is perceived a “white.”

Results

To get a first look at the data, we “stacked” the data so that each target for each participant represented a separate row (i.e., target-participant data is the level of analysis), yielding a dataset of 1,440 cases (90 participants x 16 targets). Figures 1 and 2 display all rated targets in ascending order in terms of how positively (versus negatively) they were rated (Figure 1) and how upset participants reported they would be if the target was physically harmed (Figure 2).⁵ Although not systematic, some interesting similarities and differences are evident across the figures. Across both variables, more right-leaning targets tend to score low and more left-leaning targets tend to score high, which is unsurprising given our left-leaning sample, but nonetheless seems to provide some initial support for the pervasive role of perceived political leaning of the target. For example, Darth Vader (Star Wars) seems rated more positively than right-wing figures like Mitch McConnell, Tomi Lahren, and the NRA. Participants would be more upset if Kanye West was physically harmed than if Tucker Carlson was. On the other side of the scales, Bernie Sanders and the #MeToo movement are rated more positively than Jesus Christ, and participants would be more upset if harm came to Alexandria Ocasio-Cortez (AOC) than to the Dalai Lama.

[FIGURES 1 & 2 ABOUT HERE]

Importantly, aside from the similar role of political leaning across variables, the targets do not appear in exactly the same order across the figures, suggesting some distinction in the processes underlying the types of ratings. For example, despite ranking highly on positivity, participants would be relatively less upset if Luke Skywalker were physically harmed. Further,

⁵ Each individual target was rated by between 2 and 17 participants ($M = 9$, $SD = 4$). Targets omitted from figures if they were not rated by at least 2 participants.

right-wing targets like Tomi Lahren and #AllLivesMatter ranked near the bottom in terms of positivity, but participants are relatively less likely to want to see them physically harmed. Although interestingly, these rankings do not necessarily tell us much in terms of the systematic factors that predict ratings of the different targets.

To identify the factors more predictive of positivity toward the targets as well as how upset one would be if the target were harmed, we used linear regression with robust standard errors clustered by participant to regress each of our dependent variables (positivity and ‘upset if harmed’) on participants’ ratings for the target’s race (as coded from the pre-test), warmth, intelligence, violence, and ideological distance (calculated as how conservative the target was perceived as being if the participant identified as ‘Liberal’ or ‘Moderate, leaning liberal’ (participants: $n = 63$; trials: $n = 1,008$), how liberal the target was if the participant identified as ‘Conservative’ or ‘Moderate, leaning conservative’ (participants: $n = 14$; trials: $n = 224$), and how ideologically extreme in either direction the target was perceived as being if the participant identified as ‘Moderate’ (participants: $n = 13$; trials: $n = 208$)), as well as a binary variable indicating whether the target was an individual or a group (reference category = group). Results are illustrated in Figure 3 in aggregate as well as split by participant ideology.⁶

[FIGURE 3 ABOUT HERE]

Results for both dependent variables were similar. First, individuals were rated more negatively on both dependent variables than groups. We did not have hypotheses for this variable, but the effect was robust. Contrary to either race-related hypothesis, how White the target was did not matter for positivity ($b = -.024$, $SE = .031$, $p = .440$) or being upset if the target

⁶ Results disaggregated by participant ideology are shown given the political nature of many of the targets but results for conservatives and moderates should be taken with caution given the low number of participants in each category.

was harmed ($b = -.003$, $SE = .032$, $p = .922$).⁷ Perceived warmth and intelligence were both associated with greater positivity (warmth: $b = .463$, $SE = .050$, $p < .001$; intelligence: $b = .187$, $SE = .039$, $p < .001$) and being upset if the target was harmed (warmth: $b = .291$, $SE = .060$, $p < .001$; intelligence: $b = .111$, $SE = .044$, $p < .05$), although the effects of warmth were a bit stronger than those of intelligence. Contrary to the *SCM Hypothesis*, further analysis revealed these variables did not interact (for positivity: $b = .013$, $SE = .025$, $p = .611$; for ‘upset if harmed’: $b = .039$, $SE = .030$, $p = .194$), but the main effect of intelligence suggests this form of formidability is associated with more positive evaluations and greater levels of being upset if the target is harmed, supporting the *Pray on the Weak Hypothesis* (although we would not expect feelings of positivity to correspond with tolerance of physical harm in this sense, as the motivation to avoid/escape is not necessarily indicative of positive feelings (if anything, it is the contrary)). Greater ideological distance was associated with decreased positivity ($b = -.172$, $SE = .040$, $p < .001$) and being less upset if the target was harmed ($b = -.159$, $SE = .049$, $p < .01$), supporting the *Political Alignment Hypothesis*. Finally, perceived violence of the target was associated with decreased positivity ($b = -.278$, $SE = .046$, $p < .001$) and being less upset if the target was harmed ($b = -.416$, $SE = .048$, $p < .001$), supporting the *Violence Hypothesis*.

The most evident difference across our dependent variables was the relative effects of warmth versus violence of the target. When predicting positivity, the perceived warmth of the target matters most, but when predicting being upset if the target was harmed, the perceived

⁷ The results split by participant party suggest race may matter in opposite directions when predicting positivity, albeit weakly, for liberals (feel more negatively toward White targets) versus moderates (feel more positively toward White targets); and it may not matter regardless of party for ‘upset if harmed.’ However, again, our limited sample of moderates and conservatives means we should take these results cautiously.

violence of the target matters most.⁸ It is also notable that just these 5 variables explain an unprecedented amount of variance in these dependent variables: $R^2 = .638$ when predicting positivity, and $R^2 = .536$ when predicting ‘upset if harmed’.

Conjoint Rating Task

Data

The sample for the conjoint rating task was the same as the one used for the RWTR task.

Protocol and Experimental Design

Participants engaged in a 30-trial conjoint task in which they chose between two groups (“Group A” and “Group B”) based on the following question: “Which group would you be more likely to support violence against in the form of a member of that group being punched in the face while in public?” Before the task began, participants read instructions for the task and descriptions of each of the attributes. We manipulated 8 attributes as shown in Figure 4, which is the exact table participants saw during the instructions. However, due to a technical error, for the first 71 participants, our data did not contain an indicator for which category they were in for “% of country that supports them” or “% with college degree,” and so those attributes are not contained in our analyses.⁹ The attribute order was randomized across participants in line with existing recommendations for conjoint analyses (Hainmueller, Hopkins, & Yamamoto, 2014). Although 30 is somewhat more than the number of trials used in many conjoint experiments,

⁸ We also tested whether any of the effects of race, warmth, intelligence, ideological distance, and violence depended on whether the target was a group or an individual. Only one of these interactions was significant. The positive effect of intelligence predicting positivity was stronger for groups than for individuals. We did not have any expectations for this interaction, though, and so do not discuss it further.

⁹ Analyses with the 19 participants who did have data for those variables indicated neither variable had a significant effect, but such analyses should be taken with extreme caution given the low sample size.

research has shown limited satisficing at high numbers of trials (including up to 30; Bansak, Hainmueller, Hopkins, & Yamamoto, 2018).

[FIGURE 4 ABOUT HERE]

The attributes had the following trait levels. *Race* consisted of 4 possible categories and was distributed such that the probability that a group was in a category was reflective of the racial/ethnic composition of the United States. Specifically, groups had a 60% chance of being “White,” a 15% chance of being either “African American” or “Hispanic/Latinx,” and a 10% chance of being “Asian American.” *PID* was randomly set as either “Mostly Democrats” or “Mostly Republicans.” *Income* consisted of 4 possible categories: “< \$40,000/yr”, “\$40,000/yr to \$70,000/yr”, “\$70,000/yr to \$100,000/yr”, and “> \$100,000/yr.”

Worldview consisted of 4 possible categories: conservative-authority, conservative-dominance, liberal-authority, and liberal-dominance. We included dominance and authority separately given the existing literature suggesting these may be distinct drivers of political ideology (e.g., Duckitt & Sibley, 2009). This attribute included fictional quotes said to be drawn from the social media accounts of members of the group. The quotes were constructed to reflect liberal or conservative values related to dominance and authority. As such, liberal-authority quotes included statements such as, “Tradition is overrated. We are fighting for a new and different tomorrow,” whereas conservative-authority quotes included statements such as, “We want the chaos and disorder stopped, and structure returned.” Additionally, liberal-dominance quotes included, “All we are fighting for is equality and equal opportunity,” whereas conservative-dominance quotes included, “If some people are able to use their strength and skills to dominate others, good for them.” *Violence* contained quotes supposedly taken from social media accounts as well, and consisted of “violent” quotes (e.g., “Just look at the movies. Even

superheroes are violent sometimes when it is against evil”) and “nonviolent” quotes (e.g., “We must fight with our words and votes, not fists and guns”). See the Appendix for a full list of quotes for both *worldview* and *violence*.

Description was a single descriptive word chosen randomly from one of two lists: one list that contained threat-related words ("power-hungry", "dangerous", "predators", "aggressive", "threatening") and one that contained disgust-related words ("vermin", "rats", "grotesque", "disgusting", "vile"). All quotes for the *worldview* and *violence* attributes and words for the *description* attribute were selected based on the same pre-test done for the RWTR task. Pre-test respondents rated 60 quotes constructed to reflect liberal-authority, conservative-authority, liberal-dominance, conservative-dominance, violent, or nonviolent sentiment. Quotes were rated on how much they promoted violence, dominance (versus equality), and obedience to authority. Three quotes for each category were chosen based on which quotes were rated most purely as their respective dimension. Pre-test respondents also rated 20 descriptive words according to how much they reflected being disgusting, threatening, strong (versus weak), and White (versus Non-White). Five words each for the “disgusting” and “threatening” *description* attribute were chosen based on words that most purely represented their respective category and were relatively balanced in terms of how “White” they were perceived as being. Threatening words were rated as more powerful than disgusting words, but that fits in line with the theory behind what should differentiate these dimensions according to the intergroup emotions literature. See the Appendix for a description of the pre-test.¹⁰

¹⁰ The development of the conjoint task was aided tremendously by the instructional pieces developed by Thomas Leeper (<https://github.com/leeper/conjoint-example>) and Kyle Dropp (<http://www.weebly.com/uploads/1/2/0/9/12094568/conjoint.pdf>).

Results

To analyze these data, we followed the guidance of previous work on conjoint experiments in political science (e.g., Hainmueller, Hopkins, & Yamamoto, 2014). Specifically, we “stacked” the data so that each row represented a group from a trial, and so each group in each trial was represented separately, along with separate variables for each group’s values for each attribute. The resulting data set had 5,400 rows (2 groups x 30 trials x 90 participants). We then used linear regression with robust standard errors clustered by participant to regress the binary variable of whether that group was chosen (to be punched) on the *PID*, *race*, *income*, *description*, *violence*, and *worldview* variables described above. The resulting coefficients give us the Average Marginal Component Effect (AMCE) for each attribute.

Figure 5 contains a coefficient plot of the AMCEs for the conjoint rating task in the aggregate as well as split by participant party.¹¹ We focus just on the aggregate results here. First, with regard to race, groups labeled as African American or Hispanic/Latinx were significantly *less* likely to be chosen than groups labeled as White, providing some support for the *Egalitarian Norms Hypothesis*. However, Asian American groups were no more or less likely to be chosen than White groups. Income of the group did not matter, counter to any expectations about formidability or competence.¹² Descriptions of the groups using disgusting or threatening descriptors did not have a significant effect, counter to expectations based on the intergroup emotions literature. In terms of political alignment, the broad *Political Alignment Hypothesis* is

¹¹ Here, just as with the RWTR task, caution is advised when interpreting the results for Independents and Republicans given the low number of them in the sample. Further, additional analyses involving assessments of sensitivity to which group is the reference category would be necessary to fully explicate differences across participant party (Leeper, Hobolt, & Tilley, 2020).

¹² We also modeled interactions between income and PID, worldview, and race, as it might be expected given SCM literature that the effects of perceived competence depend on the ingroup-outgroup status of the group being evaluated. None of these interactions were significant.

supported as both partisanship of the group and ideological values have significant positive effects on the probability of being chosen. However, the effect of ideological values is somewhat stronger, giving a slight edge to the *Ideological Values Hypothesis* over the *Partisanship Hypothesis*. Further, although we did not have expectations for the type of values that would matter, dominance-related values are clearly the ones that matter, and authority-related values have no effect. Finally, the *Violence Hypothesis* is strongly supported, as perceiving the group as violent is the strongest predictor of being chosen. No hypotheses related to disgust or perceived competence or formidability are supported in these analyses.¹³

[FIGURE 5 ABOUT HERE]

Discussion

What characteristics might political individuals or groups have that make them susceptible to being targets of violence? Across two experimental tasks, we found consistent evidence supporting two factors predicting willingness to condone violence against outgroups: political alignment and perceived violence. With regard to political alignment, we found support for the roles of both sheer party labels and expressed ideological values, although ideological values mattered slightly more. The effects of expressed ideological values support the idea that conflict in perceived worldviews matters above and beyond just partisan labels, in line with the Ideological Conflict Hypothesis (Brandt et al., 2014). This also supports the idea that partisan moral disengagement tends to allow for rationalization of harm against political opponents (Cassese, 2019; Kalmoe & Mason, 2019). This is especially interesting because it suggests that support for political violence, and perhaps affective polarization more broadly, reflects more than

¹³ All data, scripts, and materials can be found online via the Open Science Framework (<https://osf.io/vga2r/>).

just people butting heads with one another because they are on different “teams.” Parties reflect distinct values and worldviews, and those seem to matter independently and potentially more strongly than partisan labels alone.

Dominance-related values provoked support for violence, but authority-related values did not. Specifically, the groups most likely to be punched were those who expressed values related to "conservative dominance" (as opposed to liberal values or conservative authority values). While keeping in mind that we had a left-leaning sample, these results suggest that although sheer ingroup-outgroup dynamics are at play here, perceiving the target group as promoting a worldview/values/moral code inconsistent with yours (at least in the domain of "dominance") is more powerful. This makes intuitive sense if we think of the prototypical example of “punching Nazis.” Although authoritarian values are core to Nazi ideology, the idea that Nazis promote oppression of others (and thus encourage an illiberal, undemocratic worldview) is often what is used to justify foregoing civility with them.

We found support for independent effects of partisan labels as well. Previous research has argued that partisanship is a social identity and may be crucial in the realm of intergroup violence (e.g., Kalmoe & Mason, 2019; Iyengar & Westwood, 2015; Harel, Maoz, & Halperin, 2020; Matherus, Martinez, & Theodoridis, 2019; Pacilli et al., 2016). Independent of expressed ideological values, we found that sheer party labels mattered too. As such, a nontrivial element of support for political violence seems to be the ingroup-outgroup dynamics at play when we know that the target group is on “the other team.”

The effects of perceived violence were robust and were shown to be particularly important when it comes to attitudes toward physical harm against the group or individual (as opposed to just feeling positively or negatively toward them). In other words, individuals in our

sample were more likely to support violence against groups perceived as violent or supporting violence themselves. Although this finding may seem overly intuitive, it is worth emphasizing given that it is often overlooked in existing literature. Other research has shown the importance of “violence-justifying ideologies” in making violent extremism possible (e.g., Kruglanski et al., 2014). Further, when the “line of civility,” or at least the boundary of a peaceful democratic process, has been breached, it becomes a “no holds barred” match. If a group or individual is perceived as being violent or promoting violence, it seems to be the single most important factor (at least of those we examined) in shifting from “dislike” to acceptance of violent action.

We found somewhat mixed support for other main variables of interest. When it comes to race, our study revealed support for an *Egalitarian Norms Hypothesis* over a *Straightforward Prejudice Hypothesis*. That is, when groups were explicitly identified as African American or Latinx, people were less likely to support punching them. This seems to support theory in much political psychology work that overt racial cues lead people to inhibit racial biases due to the power of norms against racism (Mendelberg, 2001; Valentino, Hutchings, & White, 2002). This is particularly interesting given some work suggesting this over-covert distinction between racial cues has been diminished in recent politics due to the oversaturation of racial appeals in politics (Valentino, Neuner, & Vandenbroek, 2018). In our study, we did not manipulate covert versus overt appeals, but there does seem to be some reluctance to support violence against African American and Latinx groups, at least outwardly. Yet this finding was not evident in the RWTR task, in which how White the target was did not matter, which could be due to the different (and more limited) measure of race (just gauging how White the target was perceived as being in the pre-test), but it could also be that implicit-explicit dynamics were at play here, as race of the target was never overtly asked about in the main study. Perhaps the lack of an overt mention of

race diminished the buffer that egalitarian norms would otherwise provide. However, non-White targets were not more likely to be evaluated negatively either, and so we never find support for our *Straightforward Prejudice Hypothesis*.

It is important to note that even when race did matter, the effect did not exist for *all* ethnic minority groups. Asian American groups were met with more support for being punched than other racial minorities in our study, and so while egalitarian norms have the capacity to exert powerful inhibition effects, they only go so far, and perhaps only certain groups (such as African Americans or Latinx) are protected by them. This makes sense given recent rises in anti-Asian American hate crimes. There was an approximate 150% increase in anti-Asian American hate crimes in 2020 alone (California State University's Center for the Study of Hate and Extremism, 2021). These results suggest that we should take a greater rigor when looking at views of Asian Americans, and other racial and ethnic groups, in future research. Are different racial groups equally protected against by the norm of egalitarianism?

We also found some support for the role of perceived competence, although not in the way expected from the SCM or BIAS Map. Instead, in line with our *Pray on the Weak Hypothesis*, both positivity and being upset if the target were physically harmed were positively related to perceived target intelligence in the RWTR task. However, income, which we argue should also reflect competence, had no effects in the conjoint analysis. Further, if the mechanism of the effects of perceived intelligence in the RWTR task were really about fear triggering avoidance motivations, it is unclear why the effect would exist for positivity as well, and so a more intuitive explanation may simply be that intelligence was rated in a way that was similar to warmth (indicating just positive feelings toward the group), yet distinct enough to

exhibit independent effects. Future research should investigate the role of competence and formidability further.

Our study also highlights that there were differences between ratings between individuals and groups in the RWTR task. For example, in our fairly liberal sample, “Republicans” were rated quite centrally for measures such as “average positivity” and “upset if harmed.” However, individual members of the Republican Party, such as Mitch McConnell and Donald Trump, were rated extremely low for “average positivity” and “upset if harmed.” We did not have expectations for this variable. However, it may be quite important. To our knowledge, most work on affective polarization and political tolerance does not look at individuals. As such, existing literature may underestimate political intolerance and support for violence that may be targeted at individuals rather than groups. It may be beneficial to parse out individuals from groups in this field of research to better understand why this difference exists.

There are some potentially important implications of these findings regarding how we might not only predict, but also prevent groups from being the target of violence. Our results suggest that voicing support for violence against a group may, in a cyclical way (and somewhat ironic if the violence is being called for against a group perceived as being violent), actually increase the probability of violence expressed *by* the target group *against* one’s own group. If violence is seen as a justification tool – a means to an end – then “fighting fire with fire” will lead to further acts of retaliatory violence, a phenomenon not at all novel to scholars of international and intrastate conflict (e.g., Liberman, 2006).

References

- Aarøe, L., Petersen, M. B., & Arceneaux, K. (2017). The behavioral immune system shapes political intuitions: Why and how individual differences in disgust sensitivity underlie opposition to immigration. *American Political Science Review*, 111(2), 277.
- Allport, Gordon G. (1954). *The Nature of Prejudice*. New York: Perseus Books.
- Bansak, K., Hainmueller, J., Hopkins, D. J., & Yamamoto, T. (2018). The number of choice tasks and survey satisficing in conjoint experiments. *Political Analysis*, 26(1), 112-119.
- Berkowitz, L. (2012). A different view of anger: The cognitive-neoassociation conception of the relation of anger to aggression. *Aggressive behavior*, 38(4), 322-333.
- Bilewicz, M., & Vollhardt, J. R. (2012). Evil transformations: Social-psychological processes underlying genocide and mass killing. *Social psychology of social problems: The intergroup Context*, 280.
- Brandt, M. J., Reyna, C., Chambers, J. R., Crawford, J. T., & Wetherell, G. (2014). The ideological-conflict hypothesis: Intolerance among both liberals and conservatives. *Current Directions in Psychological Science*, 23(1), 27-34.
- Chapman, H. A., Kim, D. A., Susskind, J. M., & Anderson, A. K. (2009). In bad taste: Evidence for the oral origins of moral disgust. *Science*, 323(5918), 1222-1226.
- Chapman, H. A., & Anderson, A. K. (2013). Things rank and gross in nature: a review and synthesis of moral disgust. *Psychological bulletin*, 139(2), 30
- Clifford, S. (2020). Compassionate democrats and tough republicans: how ideology shapes partisan stereotypes. *Political Behavior*, 42(4), 1269-1293.

- Cottrell, C. A., & Neuberg, S. L. (2005). Different emotional reactions to different groups: a sociofunctional threat-based approach to "prejudice". *Journal of personality and social psychology*, 88(5), 770.
- Crandall, C. S., & Eshleman, A. (2003). A justification-suppression model of the expression and experience of prejudice. *Psychological bulletin*, 129(3), 414.
- Crawford, J. T., Modri, S. A. & Motyl, M. (2013) Bleeding-heart liberals and hardhearted conservatives: Subtle political dehumanization through differential attributions of human nature and human uniqueness traits. *Journal of Social and Political Psychology*, 1(1):86–104
- Cuddy, A. J., Fiske, S. T., & Glick, P. (2007). The BIAS map: behaviors from intergroup affect and stereotypes. *Journal of personality and social psychology*, 92(4), 631.
- Cuddy, A. J., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in experimental social psychology*, 40, 61-149.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of personality and social psychology*, 56(1), 5.
- Druckman, J. N., Klar, S., Krupnikov, Y., Levendusky, M., & Ryan, J. B. (2021). Affective polarization, local contexts and public opinion in America. *Nature Human Behaviour*, 5(1), 28-38.
- Druckman, J. N., & Levendusky, M. S. (2019). What do we measure when we measure affective polarization?. *Public Opinion Quarterly*, 83(1), 114-122.
- Duckitt, J., & Sibley, C. G. (2009). A dual-process motivational model of ideology, politics, and prejudice. *Psychological Inquiry*, 20(2-3), 98-109.

- Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin*, 23(3), 316-326.
- Farivar, M. (2021, March 2). Hate Crimes Targeting Asian Americans Spiked by 150% in Major US Cities. Retrieved from <https://www.voanews.com/usa/race-america/hate-crimes-targeting-asian-americans-spiked-150-major-us-cities>
- Fielding, D., & Shortland, A. (2010). 'An eye for an eye, a tooth for a tooth': Political violence and counter-insurgency in Egypt. *Journal of Peace Research*, 47(4), 433-447.
- Fiske, S. T. (2012). Warmth and competence: Stereotype content issues for clinicians and researchers. *Canadian Psychology/Psychologie Canadienne*, 53(1), 14.
- Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *Journal of personality and social psychology*, 82(6), 878
- Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in cognitive sciences*, 11(2), 77-83.
- Gilens, M., Glaser, J., & Mendelberg, T. (2001). Having a say: Political efficacy in the context of direct democracy. In *annual meeting of the American Political Science Association, San Francisco*.
- Greco, L. M., Whitson, J. A., O'Boyle, E. H., Wang, C. S., & Kim, J. (2019). An eye for an eye? A meta-analysis of negative reciprocity in organizations. *Journal of Applied Psychology*, 104(9), 1117.
- Hainmueller, J., Hopkins, D. J., & Yamamoto, T. (2014). Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political Analysis*, 22(1), 1-30.

- Harel, T. O., Maoz, I., & Halperin, E. (2020). A conflict within a conflict: intragroup ideological polarization and intergroup intractable conflict. *Current Opinion in Behavioral Sciences*, 34, 52-57.
- Harff, B. (2003). No lessons learned from the Holocaust? Assessing risks of genocide and political mass murder since 1955. *American Political Science Review*, 57-73.
- Harris, L. T., & Fiske, S. T. (2006). Dehumanizing the lowest of the low: Neuroimaging responses to extreme out-groups. *Psychological science*, 17(10), 847-853.
- Hayes, B. C., & McAllister, I. (2001). Sowing dragon's teeth: Public support for political violence and paramilitarism in Northern Ireland. *Political Studies*, 49(5), 901-922.
- Henry, P. J., Sidanius, J., Levin, S., & Pratto, F. (2005). Social dominance orientation, authoritarianism, and support for intergroup violence between the Middle East and America. *Political Psychology*, 26(4), 569-584
- Ho, A. K., Sidanius, J., Levin, D. T., & Banaji, M. R. (2011). Evidence for hypodescent and racial hierarchy in the categorization and perception of biracial individuals. *Journal of Personality and Social Psychology*, 100(3), 492.
- Horowitz, D. L. (2000). *Ethnic groups in conflict, updated edition with a new preface*. Univ of California Press.
- Huddy, L., Mason, L., & Aarøe, L. (2015). Expressive partisanship: Campaign involvement, political emotion, and partisan identity. *American Political Science Review*, 1-17.
- Inbar, Y., Pizarro, D., Iyer, R., & Haidt, J. (2012). Disgust sensitivity, political conservatism, and voting. *Social Psychological and Personality Science*, 3(5), 537-544.
- Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, not ideology a social identity perspective on polarization. *Public opinion quarterly*, 76(3), 405-431.

- Iyengar, S., & Westwood, S. J. (2015). Fear and loathing across party lines: New evidence on group polarization. *American Journal of Political Science*, 59(3), 690-707.
- Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, 22, 129-146.
- Jasko, K., Webber, D., Kruglanski, A. W., Gelfand, M., Taufiqurrohman, M., Hettiarachchi, M., & Gunaratna, R. (2020). Social context moderates the effects of quest for significance on violent extremism. *Journal of personality and social psychology*, 118(6), 1165.
- Kalmoe, N. P., & Mason, L. (2019). Lethal mass partisanship: Prevalence, correlates, & electoral contingencies. In *National Capital Area Political Science Association American Politics Meeting*.
- Kam, C. D., & Estes, B. A. (2016). Disgust sensitivity and public demand for protection. *The Journal of Politics*, 78(2), 481-496.
- Kelley, H. H. (1967). Attribution theory in social psychology. In *Nebraska symposium on motivation*. University of Nebraska Press.
- Khanna, N. (2010). "If you're half black, you're just black": Reflected appraisals and the persistence of the one-drop rule. *The Sociological Quarterly*, 51(1), 96-121.
- Klar, S., & Krupnikov, Y. (2016). *Independent politics*. Cambridge University Press.
- Klar, S., Krupnikov, Y., & Ryan, J. B. (2018). Affective polarization or partisan disdain? Untangling a dislike for the opposing party from a dislike of partisanship. *Public Opinion Quarterly*, 82(2), 379-390.

- Kruglanski, A. W., Gelfand, M. J., Bélanger, J. J., Sheveland, A., Hetiarachchi, M., & Gunaratna, R. (2014). The psychology of radicalization and deradicalization: How significance quest impacts violent extremism. *Political Psychology, 35*, 69-93.
- LeDoux, J. (1998). Fear and the brain: where have we been, and where are we going?. *Biological psychiatry, 44*(12), 1229-1238.
- Leeper, T. J., Hobolt, S. B., & Tilley, J. (2020). Measuring subgroup preferences in conjoint experiments. *Political Analysis, 28*(2), 207-221.
- Lelkes, Y., & Westwood, S. J. (2017). The limits of partisan prejudice. *The Journal of Politics, 79*(2), 485-501.
- Liberman, P. (2006). An eye for an eye: Public support for war against evildoers. *International Organization, 68*7-722.
- Lopez, G. (2017). The debate over punching white nationalist Richard Spencer in the face, explained. Retrieved from <https://www.vox.com/identities/2017/1/26/14369388/richard-spencer-punched-alt-right-trump>
- Lupton, R. N., Smallpage, S. M., & Enders, A. M. (2020). Values and political predispositions in the age of polarization: Examining the relationship between partisanship and ideology in the United States, 1988–2012. *British Journal of Political Science, 50*(1), 241-260.
- Mackie, D. M., Devos, T., & Smith, E. R. (2000). Intergroup emotions: Explaining offensive action tendencies in an intergroup context. *Journal of personality and social psychology, 79*(4), 602.
- Marcus, G. E., Theiss-Morse, E., Sullivan, J. L., & Wood, S. L. (1995). *With malice toward some: How people make civil liberties judgments*. Cambridge University Press.

- Mason, L. (2016). A cross-cutting calm: How social sorting drives affective polarization. *Public Opinion Quarterly*, 80(S1), 351-377.
- McClosky, H., & Brill, A. (1983). *The dimensions of tolerance: What Americans believe about civil liberties*. Russell Sage Foundation.
- Murphy, P. P. (2017, January 21). White nationalist Richard Spencer punched during interview CNN Politics. Retrieved from <https://www.cnn.com/2017/01/20/politics/white-nationalist-richard-spencer-punched>
- Oaten, M., Stevenson, R. J., & Case, T. I. (2009). Disgust as a disease-avoidance mechanism. *Psychological Bulletin*, 135(2), 303–321. <https://doi.org/10.1037/a0014823>
- Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of personality and social psychology*, 75(3), 811.
- Posner, D. N. (2004). Measuring ethnic fractionalization in Africa. *American journal of political science*, 48(4), 849-863.
- Rogowski, J. C., & Sutherland, J. L. (2016). How ideology fuels affective polarization. *Political Behavior*, 38(2), 485-508
- Rozin, P., Haidt, J., & McCauley, C. R. (2008). *Disgust*. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (p. 757–776). The Guilford Press.
- Tindale, R. S., Munier, C., Wasserman, M., & Smith, C. M. (2002). Group processes and the holocaust. *Understanding genocide: The social psychology of the Holocaust*, 143-161
- Tybur, J. M., Merriman, L. A., Hooper, A. E. C., McDonald, M. M., & Navarrete, C. D. (2010). Extending the behavioral immune system to political psychology: Are political conservatism and disgust sensitivity really related?. *Evolutionary Psychology*, 8(4), 147470491000800406.

- Valentino, N., Hutchings, V., & White, I. (2002). Cues That Matter: How Political Ads Prime Racial Attitudes during Campaigns. *The American Political Science Review*, 96(1), 75-90.
- Valentino, N. A., Neuner, F. G., & Vandebroek, L. M. (2018). The changing norms of racial political rhetoric and the end of racial priming. *The Journal of Politics*, 80(3), 757-771.
- Varshney, A. (2003). *Ethnic conflict and civic life: Hindus and Muslims in India*. Yale University Press.
- Westwood, S. J., Peterson, E., & Lelkes, Y. (2019). Are there still limits on partisan prejudice? *Public Opinion Quarterly*, 83(3), 584-597.
- Wilkinson, S. I. (2006). *Votes and violence: Electoral competition and ethnic riots in India*. Cambridge University Press.
- Wilson, J. P., Hugenberg, K., & Rule, N. O. (2017). Racial bias in judgments of physical size and formidability: From size to threat. *Journal of Personality and Social Psychology*, 113(1), 59-80.
- Woodward, C. (2021, January 13). AP FACT CHECK: Trump's call to action distorted in debate. Retrieved from <https://apnews.com/article/fact-check-trump-us-capitol-remarks-221518bc174f9bc3dd6e108e653ed08d>
- Young, D. (2016, March 21). The Definition, Danger and Disease of Respectability Politics, Explained. Retrieved from <https://www.theroot.com/the-definition-danger-and-disease-of-respectability-po-1790854699>.

Figure 1: Average Positivity of RWTR Targets

Figure 2: Average 'Upset if Harmed' of RWTR Targets

Figure 3: OLS Regression Predicting Positivity and ‘Upset if Harmed’ for RWTR Targets

Figure 4: Description of Attributes for Conjoint Rating Task

% of country that supports them	Percentage of U.S. with favorable attitudes toward this group according to polls.
How others refer to them	One of the most common words used to describe this group by individuals who view the group unfavorably. These words were chosen from text analysis of social media posts.
In their own words: Views on violence	A social media post from a member of the group that is representative of their views on violence .
In their own words: Values/Worldview	A social media post from a member of the group that is representative of their values and worldview .
Majority party ID of group members	The political party that most members of the group identify with.
Majority race	The race/ethnicity of most members of the group.
% with college degree	Percentage of group members with a 4-year college degree.
Median household income	The median household income of members of this group.

Figure 5: OLS Regression Predicting Being Chosen in Conjoint Rating Task

Pre-Test

Pre-Test Data and Protocol

Three hundred and seven U.S. adults were recruited to participate in a 5-minute online survey in exchange for \$0.75 through Amazon's Mechanical Turk (MTurk) on November 6, 2020. To screen out potential bots and international respondents as data were being collected, we utilized the Qualtrics protocol outline by Burleigh, Kennedy, and Clifford (2018), and found 2 participants with problematic IP addresses. Additionally, 17 participants failed one of two attention check questions. The final sample used in analyses was 288 (55.9% male, median age = 38, 85.8% with at least a college degree, median household income = \$40,000-\$60,000, 54.4% Democrat, 29.7% Republican).

A list of 51 groups and 51 individuals was compiled, as well as a list of 60 value- or violence-related quotes and a list of 20 descriptive words. Each participant rated 5 groups, 5 individuals, 6 quotes, and 2 traits on the dimensions listed below. The items rated by each participant were randomly selected from the respective lists, and so ultimately each group was rated by between 6 and 36 respondents ($M = 23.29$, $SD = 7.25$), each individual by between 7 and 35 respondent ($M = 23.45$, $SD = 7.41$), each quote by between 14 and 34 respondents ($M = 23.87$, $SD = 4.11$). All data and lists of groups, individuals, quotes, and traits, as well as their ratings, can be found online through the Open Science Framework (<https://osf.io/vga2r/>).

Pre-Test Survey Measures

[Individual ratings]

Have you heard of this individual before?

Yes
No

Please evaluate the this individual along the following dimensions:

NOTE: Order of items was randomized

Liberal – Conservative (1 'Extremely liberal' to 5 'Extremely conservative')
White - Non-White (1 'Non-White', 3 'Somewhat diverse/mixed', 5 'White')
Prone to Violence (1 'Not at all violent' to 5 'Extremely violent')
Popular/Well-Liked (1 'Very unpopular' to 5 'Very popular')
Intelligent (1 'Very unintelligent' to 5 'Very intelligent')
Cold-Warm (1 'Very cold' to 5 'Very warm')

[Group ratings]

Same as individual ratings except with the addition of:

Large in Size/Number (1 'Very small' to 5 'Very large')

[Quotes]

Please evaluate the this quote in terms of how much it promotes the following:

NOTE: Order of items was randomized

Promoting Violence (1 'Does not promote violence at all' to 5 'Promotes violence a lot')

Promoting Equality – Dominance (1 'Promotes equality' to 5 'Promotes dominance')

Promoting Obedience to Authority (1 'Does not promote authority at all' to 5 'Promotes authority a lot')

[Traits]

Please evaluate the this trait/description of a person or group of people in terms of how much it depicts that person/group as the following:

NOTE: Order of items was randomized

White - Non-White (1 'Non-White' to 5 'White')

Weak – Strong (1 'Extremely weak' to 5 'Extremely strong')

Disgusting (1 'Not at all disgusting' to 5 'Extremely disgusting')

Threatening (1 'Not at all threatening' to 5 'Extremely threatening')

Attention Checks

RWTR Task Attention Checks

[For Individuals]

Please answer each question about the following individual:

Please just mark the following responses so that we can ensure your answers are being coded correctly:

No (to whether or not you are familiar)

Somewhat positive (to how negative or positive you feel)

A little upset (to how upset if the individual was physically harmed)

Just answer randomly for the other questions.

[For Groups]

Please answer each question about the following group:

Please just mark the following responses so that we can ensure your answers are being coded correctly:

No (to whether or not you are familiar)

Somewhat positive
A little upset

(to how negative or positive you feel)
(to how upset if the group was physically harmed)

Just answer randomly for the other questions.

Self-Report Attention Check

Finally, please mark the below option that best describes how you responded to the questions and tasks in this study. You will not be penalized for which response you choose in any way. It is very important that you are honest in answering this question so that we can make sure to analyze the data correctly. Again, your answer to this question has no impact on whether or not you receive compensation for participation, or your relationship with the researchers.

I paid close attention during the tasks and followed instructions. I also read all of the survey questions and answered honestly.

I had some trouble paying close attention the entire time, but for the most part, I paid attention to instructions and answered the questions honestly.

I paid close attention to part of the study, but for most of the study, I skimmed over without paying attention or answering honestly.

I did not pay attention to the instructions and did not answer the questions honestly.

Other _____

Conjoint Rating Task Instructions

In this section, you will be presented with two groups (Group A and Group B) side-by-side, and various characteristics of each group in the rows beneath each.

Your job is to choose between the two groups in terms of which group you would be more likely to support violence against in the form of a member of that group being punched in the face while in public.

We understand that of course, you may not support violence against members of ANY group, but for the sake of this task, we ask that you make a decision **as if you HAD TO choose one**. We understand that this may be difficult, but know that we acknowledge these are forced choices you have to make.

-----PAGE BREAK-----

You will be provided with several pieces of information about the two groups.

The information for Group A versus Group B will change across trials. Please make sure to take your time to read through the information presented about each group before making your choice.

Before we start, we will briefly describe the different types of information you will see about each group.

PLEASE MAKE SURE TO READ THESE DESCRIPTIONS, so that you know what the different information means.

-----PAGE BREAK-----

NOTE: Participants were unable to move forward in the survey until this page had been shown for at least 30 seconds.

Here are the types of information you will see about each group. **PLEASE TAKE YOUR TIME TO READ THESE.**

IMPORTANTLY, these may be shown in a different order for the actual task.

% of country that supports them	Percentage of U.S. with favorable attitudes toward this group according to polls.
How others refer to them	One of the most common words used to describe this group by individuals who view the group unfavorably. These words were chosen from text analysis of social media posts.
In their own words: Views on violence	A social media post from a member of the group that is representative of their views on violence .
In their own words: Values/Worldview	A social media post from a member of the group that is representative of their values and worldview .
Majority party ID of group members	The political party that most members of the group identify with.
Majority race	The race/ethnicity of most members of the group.
% with college degree	Percentage of group members with a 4-year college degree.
Median household income	The median household income of members of this group.

NOTE: You will not be able to move on from this page until at least 30 seconds has passed.

-----PAGE BREAK-----

The task will begin on the next page.

Again, please choose which group you would be more likely to support violence against in the form of a member of that group being punched in the face while in public.

Conjoint Rating Task Example

[Attribute order randomized across participants]

Group A		Group B
Mostly Democrats	Majority party ID of group members	Mostly Republicans
White	Majority race	White
'Some laws are used just to control people. Society must dismantle those laws.'	In their own words: Values/Worldview	'Being open-minded is important. There is not just one right way to do things.'
\$70,000/yr to \$100,000/yr	Median household income	< \$40,000/yr
'We will achieve our goals peacefully.'	In their own words: Views on violence	'It isn't politically correct to say, but sometimes there are groups of people that need to be punched.'
38	% of country that supports them	54
rats	How others refer to them	aggressive
61	% with college degree	17

Which group would you be more likely to support violence against in the form of a member of that group being punched in the face while in public?

<div>Group A</div> <div>☐</div>	<div>Group B</div> <div>☐</div>
---	---

Conjoint Rating Task Attributes and Trait Levels

% of country that supports them

***NOTE: not included in main analyses due to technical error.

[Random number between 15 and 70]

How others refer to them

Random word chosen from one of the following lists:

[Disgust]

"vermin", "rats", "grotesque", "disgusting", "vile"

[Threat]

"power-hungry", "dangerous", "predators", "aggressive", "threatening"

In their own words: Views on violence

Random quote chosen from one of the following lists:

[Violent]

"Others are violent against us, and so we will be violent back.",

"Unfortunately, violence is the only way to really change things.",

"It isn't politically correct to say, but sometimes there are groups of people that need to be punched.",

"Sometimes violence is necessary.",

"Meaningful change has never happened through peaceful means alone.",

"Just look at the movies. Even super heroes are violent sometimes when it is against evil."

[Nonviolent]

"Violence is never the answer.",

"We must fight with our words and votes, not fists and guns.",

"Resorting to violence is something we would never stoop to.",

"We will achieve our goals peacefully.",

"We condemn any violence among our members.",

"The battle of words and ideas is much more important than the battle of guns and fists."

In their own words: Values/Worldview

Random quote chosen from one of the following lists:

[Conservative-Authority]

"There are too many cooks in the kitchen these days. Listen to those in charge.",

"We want the rules followed. That is our goal.",

"We want the chaos and disorder stopped, and structure returned."

[Conservative-Dominance]

"If some people are able to use their strength and skills to dominate others, good for them.",

"The most important thing is winning, or else all is lost.",

"We spend too much time catering to those who don't deserve help."

[Liberal-Authority]

"Tradition is overrated. We are fighting for a new and different tomorrow.",

"Some laws are used just to control people. Society must dismantle those laws.",

"Being open-minded is important. There is not just one right way to do things."

[Liberal-Dominance]

"All we are fighting for is equality and equal opportunity.",
"We are all in this together.",
"For too long, we have prioritized certain groups of people."

Majority Party

Random selection from:

"Mostly Democrats"
"Mostly Republicans"

Majority race

Selected such that groups had the following probabilities of having each racial/ethnic category:

"White" (60% chance)
"African American" (15% chance)
"Hispanic/Latinx" (15% chance)
"Asian American" (10% chance)

% of group with college degree

***NOTE: not included in main analyses due to technical error.

[Random number between 5 and 80]

Median household income

Random selection from:

"< \$40,000/yr"
"\$40,000/yr to \$70,000/yr"
"\$70,000/yr to \$100,000/yr"
"> \$100,000/yr"

Appendix References

Burleigh, Tyler, Ryan Kennedy, and Scott Clifford. 2018. How to Screen Out Vps and International Respondents Using Qualtrics: A Protocol (October 12). Available at SSRN: <https://ssrn.com/abstract=3265459>.