



A threat-based hate model: How symbolic and realistic threats underlie hate and aggression[☆]

Cristhian A. Martínez^{*}, Jan-Willem van Prooijen, Paul A.M. Van Lange

Department of Experimental and Applied Psychology, Institute for Brain and Behavior (IBBA), Vrije Universiteit Amsterdam, the Netherlands

ARTICLE INFO

Keywords:

Symbolic threats
Realistic threats
Hate
Aggression

ABSTRACT

Despite growing scientific attention for hate, little is known about how perceived threats may influence hate and aggression. In four preregistered online studies ($N_{total} = 1422$), we test a threat – hate – aggression model, examining the differential effects of symbolic and realistic threats on the emergence of hate, and the associations between hate and specific aggressive behaviors, across interpersonal and intergroup hate targets. In Study 1 we specify models testing the threat – hate – aggression paths. In Studies 2 (interpersonal hate) and 3 and 4 (intergroup hate) we manipulate realistic and symbolic threat perceptions, measuring hate and aggression. Across studies, hate is better predicted by symbolic than realistic threats. Also, hate consistently predicts aggressive tendencies and hurting behaviors, and interpersonal hate mediates the relationship between symbolic threats and the two aggressive behaviors while intergroup hate mediates the relationships between symbolic (and partially realistic) threats and the two aggressive behaviors. We discuss the implications of our findings for hate, threat, and prejudice research.

Hate is often seen as a key motivator of political, ethnic, or religious intolerance around the world, and seems to play an important role in interpersonal and intergroup conflict. A comprehensive definition of hate, as well as empirical research on its antecedents, unique characteristics, and outcomes is still remarkably scarce, however (Fischer, Halperin, Canetti, & Jasini, 2018; Martínez, van Prooijen, & Van Lange, 2022). A wide range of interpersonal and intergroup threats have been proposed as triggers of hate feelings (Staub, 2011; Sternberg & Sternberg, 2008), and a variety of aggressive behaviors, ranging from social exclusion to genocide have been described as hate outcomes (e.g., Baumeister & Butz, 2005; Staub, 2005). However, neither the particular threats relevant to hate have been empirically tested, nor the associations between hate and concrete aggressive behaviors have been specified.

Recent threat-based theories have argued that people perceive qualitative different threats from different targets, leading to differential emotional reactions and behaviors (Aubé & Ric, 2019; Cottrell & Neuberg, 2005; Johnston & Glasford, 2014). The integrated threat theory (ITT), propose two broad threat categories relevant to interpersonal and intergroup conflict: *realistic threats* defined as threats to safety, goals or

resources; and *symbolic threats*, defined as threats to worldviews, values and identity (Stephan & Stephan, 2000, 2017). Although these kinds of threats have been associated with various discrete emotions such as anger, disgust or fear, their relationship with hate remains unexplored. Accordingly, the goal of the present research is to examine a functional threat-hate-aggression model across interpersonal and intergroup hate targets. Our aim is three-fold. First, we examine whether realistic or symbolic threats are more relevant to eliciting hate. Second, we examine the association between hate and three specific aggressive behaviors. And third, we examine the extent to which hate mediates the relation between the two kinds of threats (realistic and symbolic) and aggression.

1. Hate

Across a wide variety of human feelings, hate (and love) can be categorized among the more complex ones (cf., Ben-Ze'ev, 2000). The affective status of hate is still under debate, as it has been characterized interchangeably as an emotion, a sentiment, a feeling or an attitude. However, different conceptualizations agree in describing hate as a strong, intense, enduring, and destructive emotional experience,

[☆] This paper has been recommended for acceptance by Dr Ernestine Gordijn.

^{*} Corresponding author at: Department of Experimental and Applied Psychology, Institute for Brain and Behavior Amsterdam (IBBA), Vrije Universiteit Amsterdam, Van der Boechorststraat 7, 1081 BT, Room, C576 Amsterdam, the Netherlands.

E-mail address: c.a.martinez@vu.nl (C.A. Martínez).

<https://doi.org/10.1016/j.jesp.2022.104393>

Received 28 October 2021; Received in revised form 4 July 2022; Accepted 15 July 2022

Available online 28 July 2022

0022-1031/© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

intended to harm or eliminate its targets physically, socially, or symbolically (Fischer et al., 2018; Martínez et al., 2022; Opatow & McClelland, 2007; Staub, 2005; Sternberg., 2003). Hate can be experienced towards specific individuals (e.g., Aumer et al., 2016) and entire out-groups (e.g., Halperin, 2008). However, interpersonal and intergroup hate share the same appraisals, goals, and outcomes, and are significantly correlated across dimensions within-individuals (Martínez et al., 2022). Therefore, we expect little differences and aim to generalize our findings across targets. In terms of its consequences, hate has been associated with a wide variety of attitudes and behaviors including intolerance, moral exclusion, hate speech, verbal and physical aggression, violence and genocide (Chetty & Alathur, 2018; Fischer et al., 2018; Halperin, Canetti-Nisim, & Hirsch-Hoefler, 2009; Staub, 2011; Sternberg., 2003). But there is scarce empirical evidence of these associations.

In line with previous theorizing, we conceptualize hate as a distinct yet composite feeling comprising simultaneously anger, contempt, and disgust, in different degrees (Hutcherson & Gross, 2011; Matsumoto, Hwang, & Frank, 2016, 2017; Sternberg & Sternberg, 2008; Sternberg., 2003). There are various arguments supporting this notion. First, hate does not meet common criteria of discrete emotions in the sense that (a) is oriented to the longer-term; (b) does not have a unique universal facial expression; (c) is semantically and physiologically hard to distinguish from anger, contempt and disgust; and (d) it is not provoked by a pre-determined kind of situations (Fischer & Giner-Sorolla, 2016). And second, there is evidence suggesting the convergence of specific anger, contempt, and disgust dimensions, contributing independently to the emergence of hate's distinctive features (Martínez et al., 2022; Matsumoto et al., 2016, 2017; See also supplementary materials). Thus, we address hate as a feeling, understood as a long-term, higher-order, composite affective state (Roseman & Steele, 2018). At the same time, compared with other negative emotions, hate has distinctive appraisals such as perceiving the targets as dispositionally evil and unchangeable, their offenses as unfair and intentional, the perception of not being able to cope with them in the future, and the goal of eliminating them (Halperin, 2008).

2. Hate and threat

Although hate has been associated with some realistic threats (e.g., threats to life or security) (Baumeister, 1997; Beck, 2000), previous theorizing has clearly emphasized the role of threats to people's morality and self-concept on the development of hate (Baumeister & Butz, 2005; Beck, 2000; Fromm, 1992; Kucuk, 2016; Staub, 2011; Van Doorn, 2018). For example, hate has been described as a reaction to perceived moral violations (Van Bavel, Ray, & Cunningham, 2018), injustice (Kucuk, 2016; Van Doorn, 2018) the appraisal of the targets as dispositionally dangerous, immoral, or evil (Baumeister, 1997; Zeki & Romaya, 2008), and threats to values, self-esteem, and identity (Baumeister & Butz, 2005; Staub, 2005).

Consistent with this theorizing, some research suggest indeed that intergroup hate is associated with disparities in values and ideology (Halperin, 2008), cultural (rather than economic) threats (Glaser, Dixit, & Green, 2002), and a need for coherence and meaning (Elnakouri, Hubley, & McGregor, 2022). One interpretation of these findings is that hate is rooted in rather abstract, longer-term, and value/self-oriented threats, which fall within the symbolic threat category according to the ITT (Stephan & Stephan, 2000, 2017). In contrast, more immediate and concrete threats to physical safety, goals, or resources (i.e., realistic threats), triggered by targets' behaviors rather than dispositions, have been more clearly associated with proportional shorter-term emotions such as anger or fear (e.g., Halperin, 2008; Hutcherson & Gross, 2011). Therefore, considering the intensity and long-term nature of hate, as well as its particular sensitivity to threats to values and identity, we argue that hate feelings are better predicted by symbolic rather than realistic threats

3. Threat-emotion-behavior

Extensive research has revealed positive associations between threat perceptions and prejudice (e.g., Cowling, Anderson, & Ferguson, 2019; Silva, Oliveira, Dias, Pinto, & Marques, 2018), xenophobic attitudes (e.g., Esses & Hamilton, 2021; Hjern & Nagayoshi, 2011), and intolerance (Shepherd, Fasoli, Pereira, & Branscombe, 2018) towards immigrant and refugee out-groups in different societies. Although hate is also motivated by threat perceptions and may drive these phenomena, it has not been directly measured and included in these accounts. For example, perceiving a minority as a threat may give rise to xenophobic attitudes and hate feelings towards them, which in turn motivates behaviors like intolerance or aggression. Although hate is regarded as critical for understanding interpersonal (Aumer et al., 2016) and intergroup (Halperin et al., 2009) conflict, compared to other emotional and attitudinal constructs it has not been addressed yet as mediator between threat perception and aggression.

The reported emotional reactions to realistic and symbolic threat perceptions tend to be negative. Among them are for instance fear, anxiety, anger, resentment, *schadenfreude*, feelings of vulnerability, and guilt (Stephan, Ybarra, & Morrison, 2009). A number of threat-based functional models propose specific configurations in the association between targets, specific types of realistic and symbolic threats, discrete negative emotions, and behaviors. For example, there is evidence that threats to goals elicit anger followed by active harm, while contamination threats (pathogen and moral), elicit disgust followed by passive harm (Johnston & Glasford, 2014).

Furthermore, in line with our argument that as compared to realistic threats, symbolic threats evoke a stronger blend of negative emotions (and therefore potentially hate feelings), there is evidence that disgust and anger mediate the associations between symbolic threats and prejudice, social distance and political intolerance towards immigrants (Wirtz, Van Der Pligt, & Doosje, 2016). Moreover, there is evidence that Westerners experience anger (motivated by group dominance) when Muslims are perceived as an economic threat, whereas they experience anger and disgust (motivated by the protection of in-group social values and norms) when Muslims are perceived as a value threat (Matthews & Levin, 2012). This suggests, first, that emotions play an important mediating role in the associations between threat and behavior (see also Aubé & Ric, 2019; Cottrell & Neuberg, 2005). And second, that compared to realistic threats, symbolic threats elicit either different or a stronger blend of negative emotions, some of them corresponding to hate's components (Hutcherson & Gross, 2011; Matsumoto et al., 2016, 2017; Sternberg & Sternberg, 2008). Therefore, we argue that hate plays an important mediating role between symbolic threat perceptions and aggression, and we aim to test if this effect is stronger compared to realistic threats.

4. The current research

To summarize, in the present research we propose that (a) relative to realistic threats, symbolic threats are better predictors of hate; (b) hate predicts aggression; and (c) hate mediates the relation between symbolic (rather than realistic) threat perceptions and aggression. To examine these propositions, we designed four studies. In Study 1, we test a conceptual model of hate, entering symbolic and realistic threats as predictors of hate, and hate as predictor of aggressive tendencies, hurting behaviors, and social hostility. In Study 2, we conduct an experiment in which we manipulate threat perceptions (realistic versus symbolic), and test their effect on interpersonal hate and aggressive tendencies and hurting behaviors, as well as the mediation effect of interpersonal hate between the two types of threats and the two aggressive behaviors. Finally, in Studies 3 and 4, we conduct experiments to replicate Study 2 findings for intergroup hate with two different experimental manipulations. All studies' predictions, designs, and planned sampling and analyses, were preregistered and approved by the ethics board of the Vrije

Universiteit Amsterdam, and all materials, data, analyses, and pre-registrations are available on the open science framework (<https://osf.io/ksx74>).

5. Study 1

A preliminary pilot study (see supplementary material) revealed significant associations between symbolic threats and hate across interpersonal and intergroup targets, but not significant associations between realistic threats and hate, in line with predictions. In Study 1, we aim to conduct a more stringent test of these associations, as well as the association between hate and direct aggressive behaviors using Structural Equation Modeling. As preregistered, we expect positive associations between symbolic threats and hate across targets. We also expect weaker associations between realistic threats and hate, compared to the associations between symbolic threats and hate. Finally, we expect positive associations between hate and aggressive tendencies, hurting behaviors and social hostility.

5.1. Methods

5.1.1. Participants

In order to determine the sample size we conducted a series of a priori Monte Carlo simulations (Wolf, Harrington, Clark, & Miller, 2013). On the basis of previous theorizing and the results of the pilot study, we generated a hypothesized population model with the approximately expected parameter values. The model converged in 100% of the iterations using 500 replications, with a sample of 250, and alpha of 0.05, and replicated with two different seed values, showing good average fit measures: $\chi^2(10, N = 250) = 12.42$, CFI = 0.99, TLI = 0.97, RMSEA = 0.065, SRMR = 0.045, and good average power (>0.50), and coverage (range = 0.93–0.97), for the parameters of interest.

Considering the suggested $N = 250$, we collected an online US sample through Prolific ($N = 300$). We excluded 12 participants who provided incomprehensible or unrelated open answers; 1 participant who failed the two attention checks; and 1 participant who did not report hate targets but answered the measures. This rendered a sample of $N = 286$, from which 12 reported not having either interpersonal or intergroup hate targets and were therefore excluded; 266 reported interpersonal hate targets; and 267 reported intergroup hate targets. We screened for multivariate outliers using Mahalanobis distances within the interpersonal and intergroup sub-samples, and excluded 1 meaningful outlier from the interpersonal subset and 2 from the intergroup subset, obtaining final samples of $N = 265$ for the interpersonal condition, 38.1% males, 60.8% females, 1.13% other gender (age $M = 34.2$, $SD = 10.06$), and $N = 265$ for the intergroup condition, 37% males, 61.9% females, 1.13% other gender (age $M = 34.2$, $SD = 10.07$).

5.1.2. Procedure

After providing their informed consent, participants completed the outcome measures which were administered in random order. Participants were asked to describe in a short paragraph an individual that they hate (in the interpersonal condition) and a group that they hate (in the intergroup condition), and to briefly explain the reasons why they hate their targets. At the end of the session, participants completed a socio-demographic form, application control questions, and were debriefed. The session took on average 20 min to be completed, and participants were paid £1.73 for their participation.

5.1.3. Measures

Hate was measured with an adaptation of the Passionate Hate Scale (PHS) developed by Zeki and Romaya (2008). The hate scale is based in the triangular theory of hate (Sternberg & Sternberg, 2008; Sternberg, 2003), and has been used previously in experimental research by the authors. It comprises three sub-scales (1 = *strongly disagree*, to 7 = *strongly agree*), with four items each, measuring anger (e.g., I have

aggressive thoughts about this person/group); contempt (e.g., I really despise this person/group); and disgust (e.g., this person/group is really disgusting). The scale showed good reliability across interpersonal ($\alpha = 0.89$), and intergroup targets ($\alpha = 0.89$).¹

Threat perceptions were measured with an adaptation of the threat perceptions scale (Stephan & Stephan, 2017). It comprises two sub-scales measuring realistic threats with six items (e.g., this person/group blocks my goals; this person/group is a threat to my personal safety); and symbolic threats with seven items (e.g., this person/group is the opposite of what I am; this person/group has completely different values than me) (1 = *strongly disagree*, to 7 = *strongly agree*). Both realistic and symbolic threat sub-scales showed good reliability across interpersonal (realistic, $\alpha = 0.90$; symbolic, $\alpha = 0.86$) and intergroup (realistic, $\alpha = 0.87$; symbolic, $\alpha = 0.87$) targets.

The three aggressive behaviors² were measured by means of three behavioral tasks. We measured *Aggressive tendencies* using the voodoo doll paradigm (Dewall et al., 2013). In this task participants were asked to insert the number of pins they wish (from 0 to a maximum of 10) by clicking into a human figure representing their targets. We measured *Hurting behaviors* using an adaptation of the tangram help/hurt task (Saleem, Anderson, & Barlett, 2015). In this task, participants were asked to imagine that their reported hate targets would compete for a \$25 prize if they complete 11 tangram puzzles selected by the participant in 10 min. Afterwards, participants were instructed to assign the 11 tangram puzzles to their targets out of 30 possible, divided into 10 easy, 10 medium, and 10 hard. Hurtful behavior was operationalized as the number of hard tangrams assigned >1 . We measured *Social hostility* using an adapted version of the Social Mindfulness Paradigm (Van Doesum, Van Prooijen, Verburgh, & Van Lange, 2016; Van Lange & Van Doesum, 2015). In this task participants imagine to play a choice game against their hated targets. The task consisted in choosing one object out of a series that contained one unique object (e.g., a series of three umbrellas containing two blue and one yellow), across 24 trials (12 control and 12 experimental). As choosing the unique object limits the variety of options to the other player, these choices were coded as socially hostile (Van Doesum et al., 2016). The social hostility score is therefore the proportion of socially hostile choices across the experimental trials.

5.1.4. Design and analytical strategy

To analyze the data, we employed structural equation modeling using the *lavaan* package (Rosseel, 2012) for the R environment. With this model we aim to test how realistic and symbolic threat perceptions predict hate, and how hate predicts aggressive tendencies, hurting behaviors and social hostility. We conducted separate analyses by condition (interpersonal versus intergroup). Missing data were pairwise excluded from the analyses.

5.2. Results

5.2.1. Interpersonal hate

The interpersonal hate model showed an acceptable fit to the data, $\chi^2(6, N = 265) = 26.38$, $p < .001$, CFI = 0.95, TLI = 0.88, RMSEA =

¹ In the supplementary materials (Figure S5, S6; Tables S5, S6), we provide further psychometric information of the hate measurement, as well as alternative preregistered models in which hate is conceptualized as a latent variable comprising anger, contempt and disgust as indicators, and aggression as a latent variable comprising aggressive tendencies, hurting behaviors and social hostility as indicators, to control for measurement error. Results replicate entering hate as a latent variable.

² Our three aggressive behaviors' measures are intended to be indicators of different aspects of the broader concept of aggression. The *Aggressive tendencies* measure reflect a proxy of harming the targets directly; the *Hurting Behaviors* measure reflect the idea of blocking target's goals; and the *Social Hostility* measure the idea of limiting target's choices.

0.113, $SMRM = 0.040$. Results uncovered that interpersonal hate is significantly predicted by symbolic threats, ($\beta = 0.61, z = 11.40, p < .001$), and to a lesser extent by realistic threats ($\beta = 0.15, z = 2.78, p = .005$). Moreover, interpersonal hate significantly predicted aggressive tendencies, ($\beta = 0.45, z = 8.23, p < .001$), and hurting behaviors, ($\beta = 0.41, z = 7.29, p < .001$), but not social hostility, ($\beta = 0.09, z = 1.51, p = .131$). These results suggest that although both symbolic and realistic threats significantly predict hate, symbolic threats are stronger predictors of hate as expected. Moreover, hate feelings predict aggressive behavioral tendencies and the willingness to block hated targets' goals, but it does not predict subtle expressions of social hostility. Results are summarized in Fig. 1A.

5.2.2. Intergroup hate

The intergroup hate model showed an excellent fit to the data, $\chi^2(6, N = 265) = 13.12, p = .041, CFI = 0.99, TLI = 0.96, RMSEA = 0.067, SMRM = 0.036$. Results revealed that intergroup hate is significantly predicted by symbolic threats, ($\beta = 0.65, z = 9.08, p < .001$), but not by realistic threats ($\beta = -0.04, z = -0.57, p = .566$). Moreover, results revealed that intergroup hate significantly predicts aggressive tendencies, ($\beta = 0.47, z = 8.64, p < .001$), hurting behaviors, ($\beta = 0.43, z = 7.65, p < .001$), and also, albeit less strongly, social hostility, ($\beta = 0.15, z = 2.49, p = .013$). These results suggest that symbolic threats are more relevant in predicting intergroup hate than realistic threats as expected. Moreover, hate feelings predict aggressive behavioral tendencies and the willingness to block hated targets' goals, and to a lesser extent, subtle expressions of social hostility. Results are summarized in Fig. 1B.

5.3. Discussion

Study 1 suggests for both interpersonal and intergroup targets, that symbolic threats are better predictors of hate than realistic threats. Furthermore, hate predicts aggressive tendencies and blocking targets' goals across interpersonal and intergroup targets, but interpersonal hate does not predict social hostility, expressed as limiting targets' choices.

6. Study 2

Extending the correlational findings of Study 1, in Study 2 we provide an experimental test to provide evidence relevant to threat as a cause of hate and aggression. For this purpose, we manipulated threat perceptions (no threat vs. realistic threat vs. symbolic threat) while measuring hate, aggressive tendencies, and hurting behaviors as outcome variables. We excluded social hostility from studies 2, 3 and 4 because unlike the other two explicit measures of aggression, its associations with hate in Study 1 were weak for intergroup hate and not even significant for interpersonal hate. This might be explained because compared to the other two indicators of aggression social hostility is a relatively subtle form. We predict that symbolic threats will elicit stronger hate feelings than realistic threats, and that hate will predict aggression. Moreover, we predict that hate will positively mediate the effects of symbolic – but not realistic – threats on aggression.³

6.1. Methods

6.1.1. Participants

An a-priori power analysis for a one-way GLM with three conditions, and entering $\alpha = 0.05$, and $1 - \beta = 0.80$, suggested a sample size of $N =$

339, in order to detect small to medium effect sizes ($f = 0.17$). We collected an online sample of US adults whose native language was English through Prolific ($N = 400$). We excluded 2 participants with incomplete data; 2 participants who failed the attention check; and 1 participant who failed the tangram task checks, yielding a sample of $N = 395$. After screening for multivariate outliers using Mahalanobis distances, we removed 5 meaningful outliers, obtaining a final sample of $N = 390$ adult participants, 45.4% males, 53.1% females, and 1.54% other gender (age $M = 35, SD = 12.6$), with different occupations and educational levels. Sensitivity power analyses conducted with our final sample using G*power (Faul, Erdfelder, Lang, & Buchner, 2007), with $\alpha = 0.05$ and $1 - \beta = 0.80$, yielded a minimum detectable effect size of $f = 0.16$.

6.1.2. Procedure

After giving their informed consent, participants were randomly assigned to the control, realistic threats, or symbolic threats condition. After the manipulation, participants completed the measures of hate, aggressive tendencies, and hurting behaviors in a fixed order. Participants then provided basic demographics and were debriefed. The instrument took 15 min on average to complete, and participants were paid £1.50 for their participation.

6.1.3. Manipulation

We designed an “argumentation contest” in which participants were instructed to prepare a short speech in order to compete against the speech of an imaginary contender, under the criteria of the most convincing argument. Participants chose one out of three topics, indicated their agreement or disagreement with the topic, and wrote their speech in a short paragraph defending their position. After that, depending on (a) the condition, (b) the chosen topic, and (c) their agreement or disagreement, they were presented with the scores of the contender's speech and the results of the contest.

In the *realistic threats condition*, participants competed for an imaginary reward of \$100. They chose between neutral topics (i.e., education, technology, and lifestyle), and we showed them results indicating that a) the other player agreed with them about the issue, and b) the other player's speech was better, and therefore he wins the contest and the \$100 reward. In the *symbolic threats condition*, participants did not compete for any economic reward. They chose between value-laden topics (i.e., abortion, legalization of drugs, and same sex marriage), and we showed them results indicating that a) the other player completely disagreed with them about the issue, and b) the other player's speech was better, and therefore he wins the contest. Finally, in the *no threat condition*, participants did not compete for any economic reward. They chose between the same topics from the realistic threats condition, and we showed them results indicating that a) the other player agreed with them about the issue, and b) participant's speech was better, and therefore they win the contest. At the end of the manipulation, participants completed two comprehension checks, and two manipulation checks (1 = *not at all*, to 7 = *totally*): “Do you think that this kind of person pose a threat for achieving your goals?”, assessing realistic threats, and “Do you think that this kind of person pose a threat to your values?”, assessing symbolic threats.

6.1.4. Measures

Hate was measured with the same instrument as Study 1 ($\alpha = 0.94$). Aggressive tendencies were again measured with the voodoo doll task, and Hurting behaviors with the tangram task. All the measures referred to the fictional contender in the argumentation task.

6.1.5. Design and analytical strategy

We tested one-way GLM's for each of our three dependent measures using the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2017), for the R environment. Afterwards, we conducted post-hoc pairwise comparisons on the estimated marginal means using the

³ Additionally, in Studies 2 and 3 we predicted that targets perceived as symbolic threats would elicit higher levels of aggression than targets perceived as realistic threats. Since this preregistered prediction is only relevant for interpreting the mediation models, we report these results in the supplementary materials and focus here on the more relevant associations between hate and the aggression indicators.

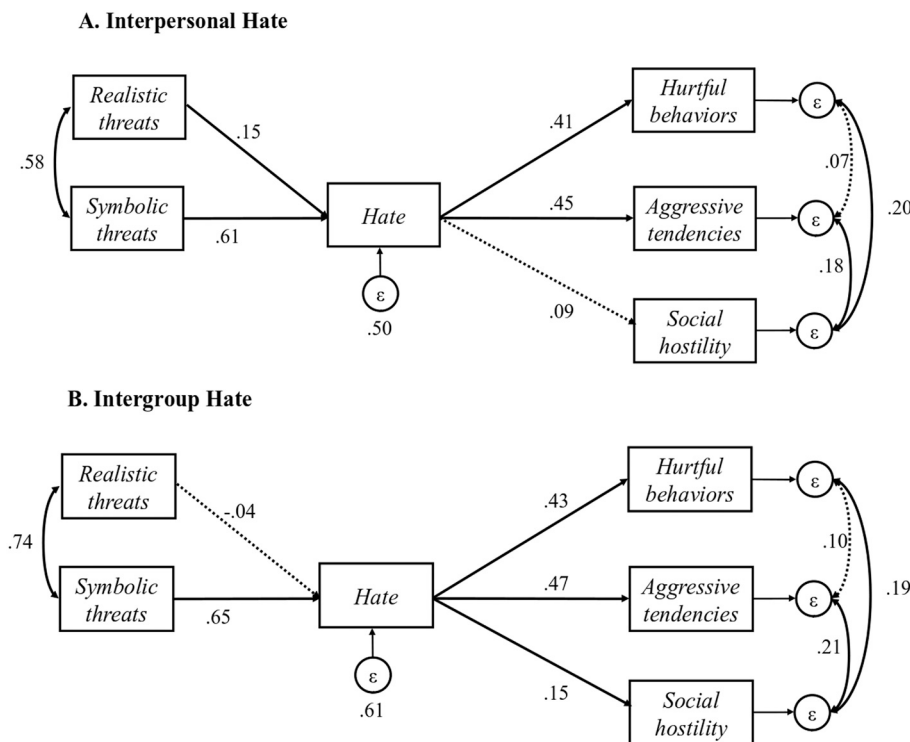


Fig. 1. Path models of the relations between threat perceptions, hate and aggression in the interpersonal (A) and the intergroup (B) target conditions. Note. Continuous arrows represent significant paths ($p < .05$). Dotted arrows represent non-significant paths. All coefficients are standardized.

emmeans package (Lenth, 2020). To test the effects of hate on aggression we conducted simple regression models. To test the mediation effects, we fitted two separate mediation models using the lavaan package (Rosseel, 2012). Following the recommendations of Hayes and Preacher (2014), we entered threat perception as a multi-categorical predictor of aggressive tendencies and hurting behaviors separately, using an indicator coding scheme. Missing data were pairwise excluded from the analyses.

6.2. Results

6.2.1. Manipulation checks

In the realistic threats condition, a paired t -test revealed that participants perceived significantly stronger threats to their goals ($M = 2.78$, $SD = 1.81$), than to their values ($M = 1.88$, $SD = 1.25$), $t(129) = 7.62$, $p < .001$, $d = 0.67$. In the symbolic threats condition, participants perceived stronger threats to their values ($M = 4.59$, $SD = 2.17$), than to their goals ($M = 4.14$, $SD = 2.00$), $t(122) = -3.25$, $p < .01$, $d = 0.29$. Finally, in the control condition, participants perceived stronger threats to their goals ($M = 2.13$, $SD = 1.36$), than to their values ($M = 1.85$, $SD = 1.26$), $t(136) = 3.38$, $p < .001$, $d = 0.29$, suggesting that participants perceived more realistic than symbolic threats in this context. These findings indicate that our manipulation was successful.⁴

6.2.2. Effects of threat perceptions on hate

Results revealed a main effect of condition on hate, $F(2, 387) = 99.78$, $p < .001$, $f = 0.72$. As predicted, post-hoc marginal means

contrasts indicated that symbolic threats elicit higher levels of hate ($M = 3.17$, $SE = 0.07$) than realistic threats ($M = 2.01$, $SE = 0.07$), $B = -1.16$, $SE = 0.10$, $t = -11.34$, $p < .001$, $d = -1.43$, 95% CI $[-1.40, -0.92]$. Results also showed that symbolic threats induce higher levels of hate, as compared to the control condition ($M = 1.85$, $SE = 0.07$), $B = -1.32$, $SE = 0.10$, $t = -13.09$, $p < .001$, $d = -1.63$, 95% CI $[-1.56, -1.08]$; and that there are no significant differences in hate scores between the realistic threats condition, and the control condition, $B = -0.16$, $SE = 0.01$, $t = -1.62$, $p = .237$, $d = -0.20$, 95% CI $[-0.40, 0.07]$. Results are summarized in Fig. 2A.

6.2.3. Associations between hate and aggression

Results revealed an effect of hate on aggressive tendencies, $F(1, 388) = 38.57$, $p < .001$, $f = 0.32$, suggesting that hate significantly predicts aggressive tendencies, $B = 0.77$, $SE = 0.12$, $t = 6.21$, $p < .001$, 95% CI $[0.53, 1.02]$. Similarly, results revealed a significant effect of hate on hurting behaviors, $F(1, 388) = 57.04$, $p < .001$, $f = 0.38$, suggesting that hate significantly predicts hurting behaviors, $B = 1.22$, $SE = 0.16$, $t = 7.55$, $p < .001$, 95% CI $[0.90, 1.53]$. Results are depicted in Fig. 2B and C.

6.2.4. Mediation effects of hate between threat perceptions and aggression

To test the mediation hypothesis, we fitted two mediation models on our aggression measures, one for aggressive tendencies and one for hurting behaviors. To calculate the parameter estimates using threat perception as the multi-categorical predictor, we coded the three conditions as a $k - 1$ variable, using the realistic threats condition as the reference group. We calculated the standard errors and 95% confidence intervals of the parameters using bootstrapping with 1000 iterations. Results are depicted in Fig. 3.

Results indicate that as compared to realistic threat perceptions, symbolic threat perceptions significantly predict hate feelings, $B = 1.16$, $SE = 0.12$, $z = 9.88$, $p < .001$, 95% CI $[0.92, 1.40]$, and with respect to the control condition, realistic threats only marginally predict hate, $B = 0.16$, $SE = 0.08$, $z = 1.93$, $p = .053$, 95% CI $[-0.004, 0.33]$.

⁴ Although the manipulations worked as expected, the levels of realistic threats perceptions are higher in the symbolic threat condition than in the realistic threat condition. To address this, we conducted additional mediation analyses with the entire sample using the manipulation checks themselves as predictors. Results replicated the mediation effects reported here (see supplementary material).

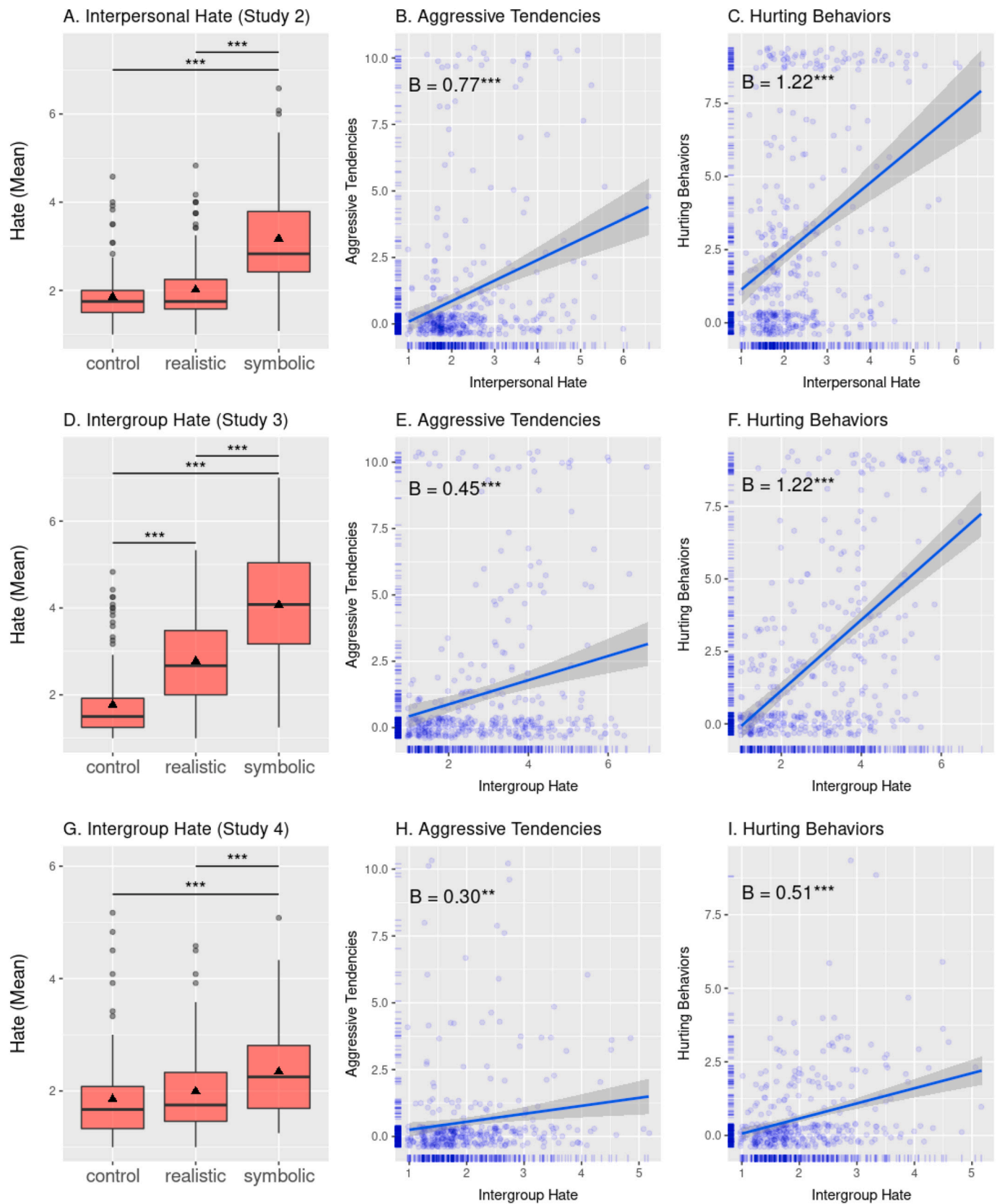
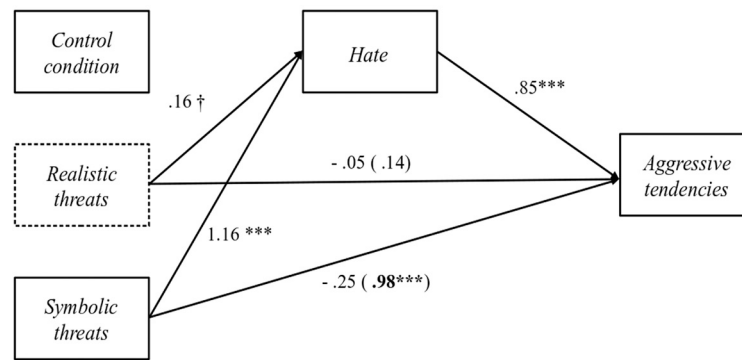


Fig. 2. Hate by condition and associations between hate, aggressive tendencies, and hurting behaviors across Studies 2 (first row), 3 (second row), and 4 (third row). Note. Triangles represent mean scores. $p < .01^{**}$, $p < .001^{***}$.

A. Aggressive tendencies



B. Hurting behaviors

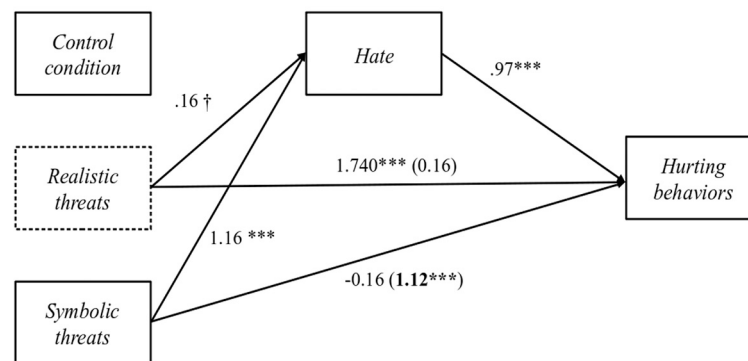


Fig. 3. Interpersonal hate mediation effects of threat perceptions on aggressive tendencies (A), and hurting behaviors (B), Study 2.

Note. Dashed boxes represent the reference condition (i.e., realistic threats) of the $k - 1$ multi-categorical predictor. Significant relative indirect effects are in bold. $p < .05^*$, $p < .01^{**}$, $p < .001^{***}$, $\dagger p = .053$.

The aggressive tendencies model (Fig. 3A), revealed that hate significantly predict aggressive tendencies, $B = 0.85$, $SE = 0.20$, $z = 4.35$, $p < .001$, 95% CI [0.51, 1.28]. In line with our predictions, both the relative direct effect of realistic threats on aggressive tendencies, $B = -0.05$, $SE = 0.27$, $z = -0.17$, $p = .862$, 95% CI [-0.58, 0.46], and the relative indirect effect of realistic threats on aggressive tendencies through hate, $B = 0.14$, $SE = 0.08$, $z = 1.64$, $p = .10$, 95% CI [0.01, 0.36], were not significant. Results also yielded a non-significant relative direct effect between symbolic threats and aggressive tendencies, $B = -0.25$, $SE = 0.41$, $z = -0.61$, $p = .545$, 95% CI [-1.02, 0.62], but in line with our predictions, results showed a significant relative indirect effect of symbolic threats on aggressive tendencies through hate, $B = 0.98$, $SE = 0.24$, $z = 4.11$, $p < .001$, 95% CI [0.56, 1.50]. Although symbolic threats do not directly translate into aggressive tendencies, we do find evidence for an indirect effect through hate.

The hurting behaviors model (Fig. 3B) yielded similar results. Hate also significantly predicted hurting behaviors, $B = 0.97$, $SE = 0.20$, $z = 4.77$, $p < .001$, 95% CI [0.58, 1.35]. Contrary to the aggressive tendencies model, results revealed a significant relative direct effect of realistic threats on hurting behaviors, $B = 1.74$, $SE = 0.35$, $z = 4.93$, $p < .001$, 95% CI [1.08, 2.48], but also a non-significant relative indirect effect of realistic threats on hurting behaviors through hate, $B = 0.16$, $SE = 0.09$, $z = 1.67$, $p = .10$, 95% CI [0.004, 0.39]. The relative direct effect between symbolic threats and hurting behaviors was nonsignificant, $B = -0.16$, $SE = 0.51$, $z = -0.31$, $p = .753$, 95% CI [-1.13, 0.90], but again, in line with our predictions we found a significant relative indirect effect of symbolic threats on hurting behaviors through hate, $B = 1.12$, $SE = 0.26$, $z = 4.22$, $p < .001$, 95% CI [0.64, 1.69]. Symbolic threats indirectly predict hurting behaviors through hate.

6.3. Discussion

Study 2 provides experimental evidence that symbolic threats, and not realistic threats, increases hate feelings, and indirectly predicts aggression (via hate). Hate seems to be a key mediator in the threat-aggression link, particularly for symbolic threats.

7. Study 3

Study 3 sought to experimentally test the effects of threat on inter-group hate, employing a similar design as Study 2. We predicted that group targets perceived as posing symbolic (rather than realistic) threats elicit higher hate feelings, and that hate predicts aggression. Moreover, we predicted that hate mediates the effects of symbolic – but not realistic – threats on aggression.

7.1. Methods

7.1.1. Participants

Following the suggested sample size from the a-priori power analysis in Study 2 (with the exact same parameters), we collected an initial online sample of $N = 400$ adults from the U.S. whose native language was English, and who did not participate previously in Studies 1 or 2, from Prolific Academic. We excluded one participant with incomplete data and one participant who failed the attention check. Hurting behavior scores from two participants who failed the Tangram task checks were coded as missing. After screening for multivariate outliers using Mahalanobis distances, we removed four meaningful outliers, obtaining a final sample of $N = 394$ adult participants, 39.6% males (age $M = 34.0$, $SD = 11.4$), 59.1% females (age $M = 35.0$, $SD = 13.0$), and

1.27% other gender (age $M = 26.2$, $SD = 5.45$), with different occupations and educational levels. Sensitivity power analyses conducted with our final sample using G*power (Faul et al., 2007), with $\alpha = 0.05$ and $1 - \beta = 0.80$, yielded a minimum detectable effect size of $f = 0.16$.

7.1.2. Procedure, design and analytical strategy

The procedure, design and analytical strategy of Study 3 were the same as Study 2, except for the manipulation (described below). The study took 14 min on average to be completed, and participants were paid £1.50 for their participation. Missing data were pairwise excluded from the analyses.

7.1.3. Manipulations and measures

Participants were instructed to choose one out of three controversial topics of their interest (i.e., abortion, legalization of drugs, and same-sex marriage) indicating their position (in favor or against) about the issue, and elaborating on their position in a short text. Next, we showed participants a description of a new (fictional) political movement in U.S., and its positions about the issue chosen by each participant. In the *symbolic threats condition*, the political party was described as (a) holding the opposite position from the participant regarding the controversial issue (e.g., if the participant is in favor of abortion, the political party was described as against abortion), and (b) as supporting paid online workers with economic incentives. In the *realistic threats condition*, the political party was described as (a) holding the same position as the participant regarding the controversial issue, and (b) as concerned about the lack of regulation of paid online work and aiming to ban it. Finally, in the *no threat (control) condition*, the political party was described as (a) holding the same position of the participant regarding the controversial issue, and (b) as supporting paid online work with economic incentives. At the end of the manipulation, participants answered two comprehension checks, and proceeded to the outcome measures. We employed the same measures from Study 2, but this time aimed towards the political party (and one hypothetical member in the aggression measures). Accordingly, participants answered the two manipulation checks; the hate scale ($\alpha = 0.95$); the aggressive tendencies measure (i.e., the voodoo doll task); and the hurting behaviors measure (i.e., the Tangram task).

7.2. Results

7.2.1. Manipulation checks

In the realistic threats condition, participants perceived more threats to their goals ($M = 4.49$, $SD = 1.87$), than to their values ($M = 3.13$, $SD = 1.85$), $t(133) = 8.72$, $p < .001$, $d = 0.75$. In the symbolic threats condition, participants perceived more threats to their values ($M = 5.40$, $SD = 1.87$), than to their goals ($M = 4.42$, $SD = 2.12$), $t(130) = -6.17$, $p < .001$, $d = -0.54$. In the control condition participants perceived little threats to their goals ($M = 1.64$, $SD = 1.35$), and values ($M = 1.67$, $SD = 1.44$), $t(128) = -0.55$, $p = .581$, $d = -0.05$. These results indicate that the manipulation was successful.

7.2.2. Effects of threat perceptions on hate

Results revealed a main effect of the threat manipulation on hate, $F(2,391) = 165.53$, $p < .001$, $f = 0.92$. As predicted, post-hoc marginal means contrasts indicated that symbolic threats elicit higher levels of hate ($M = 4.07$, $SE = 0.09$) than realistic threats ($M = 2.77$, $SE = 0.09$), $B = -1.30$, $SE = 0.13$, $t = -10.31$, $p < .001$, $d = -1.27$, 95% CI $[-1.59, -1.00]$. Symbolic threats induced higher levels of hate than the control condition ($M = 1.77$, $SE = 0.09$), $B = -2.31$, $SE = 0.13$, $t = -18.14$, $p < .001$, $d = -2.25$, 95% CI $[-2.60, -2.01]$. Also, hate scores were higher in the realistic threats condition than in the control condition, $B = -1.01$, $SE = 0.13$, $t = -7.97$, $p < .001$, $d = -0.98$, 95% CI $[-1.31, -0.71]$. Results are summarized in Fig. 2D.

7.2.3. Associations between hate and aggression

Results revealed an effect of hate on aggressive tendencies, $F(1, 392) = 21.93$, $p < .001$, $f = 0.24$, suggesting that hate significantly predicts aggressive tendencies, $B = 0.45$, $SE = 0.10$, $t = 4.68$, $p < .001$, 95% CI $[0.26, 0.65]$. Similarly, results revealed a significant effect of hate on hurting behaviors, $F(1, 392) = 175.4$, $p < .001$, $f = 0.67$, suggesting that hate significantly predicts hurting behaviors, $B = 1.22$, $SE = 0.09$, $t = 13.25$, $p < .001$, 95% CI $[1.04, 1.40]$. Results are summarized in Fig. 2E and F.

7.2.4. Mediation effects of hate between threat perceptions and aggression

Results indicate that as compared to realistic threats, symbolic threats significantly predict hate feelings, $B = 1.30$, $SE = 0.14$, $z = 9.53$, $p < .001$, 95% CI $[1.03, 1.55]$, and with respect to the control condition, realistic threats also predict hate, $B = 1.01$, $SE = 0.11$, $z = 9.10$, $p < .001$, 95% CI $[0.81, 1.26]$. The aggressive tendencies model (Fig. 4A), revealed that hate significantly predicts aggressive tendencies, $B = 0.51$, $SE = 0.19$, $z = 2.69$, $p = .006$, 95% CI $[0.14, 0.88]$. In line with our predictions, we found a non-significant relative direct effect of realistic threats on aggressive tendencies, $B = 0.06$, $SE = 0.37$, $z = 0.17$, $p = .857$, 95% CI $[-0.69, 0.74]$. However, against our predictions, the relative indirect effect of realistic threats on aggressive tendencies through hate was significant, $B = 0.51$, $SE = 0.20$, $z = 2.60$, $p < .009$, 95% CI $[0.15, 0.93]$. On the other hand, the relative direct effect between symbolic threats and aggressive tendencies was nonsignificant, $B = -0.32$, $SE = 0.45$, $z = -0.71$, $p = .487$, 95% CI $[-1.22, 0.53]$. As predicted, however, results showed a significant relative indirect effect of symbolic threats on aggressive tendencies through hate, $B = 0.66$, $SE = 0.26$, $z = 2.52$, $p = .010$, 95% CI $[0.17, 1.24]$, suggesting that symbolic threats indirectly predict aggressive tendencies through intergroup hate – a finding that is consistent with Study 2 results.

The hurting behaviors model (Fig. 4B) revealed that hate also significantly predicts hurting behaviors, $B = 1.11$, $SE = 0.13$, $z = 8.60$, $p < .001$, 95% CI $[0.86, 1.36]$. Results revealed a non-significant relative direct effect of realistic threats on hurting behaviors, $B = 0.38$, $SE = 0.29$, $z = 1.31$, $p = .194$, 95% CI $[-0.14, 1.01]$, and also a significant relative indirect effect of realistic threats on hurting behaviors through hate, $B = 1.12$, $SE = 0.20$, $z = 5.63$, $p < .001$, 95% CI $[0.77, 1.57]$. Results also yielded a non-significant relative direct effect between symbolic threats and hurting behaviors, $B = 0.19$, $SE = 0.42$, $z = 0.45$, $p = .642$, 95% CI $[-0.65, 1.05]$. In line with predictions, we found a significant relative indirect effect of symbolic threats on hurting behaviors through hate, $B = 1.44$, $SE = 0.24$, $z = 6.07$, $p < .001$, 95% CI $[1.04, 1.96]$. These findings mirror those on aggressive tendencies.

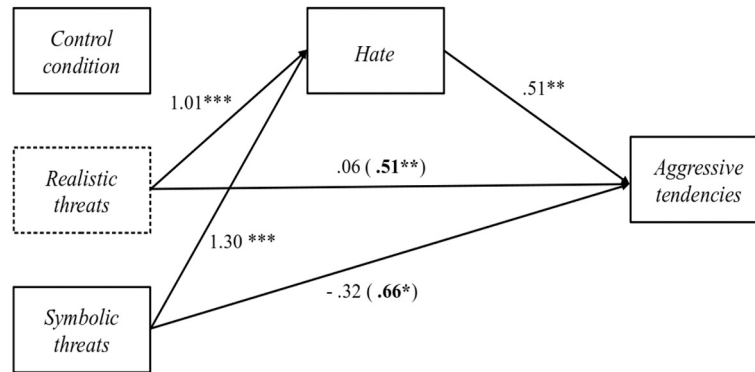
7.3. Discussion

Study 3 reveals that also at the intergroup level, symbolic (more so than realistic) threats predict hate feelings and – particularly – hurting behaviors. Unlike Study 2, results indicate that intergroup targets in the realistic threat condition elicit higher levels of hate compared to the control condition. This could explain also the mediation results, which indicated that both symbolic and realistic threats predict hate, and that hate mediates the relations between both symbolic and realistic threats and aggressive outcomes. In sum, it seems that groups posing realistic threats also predict hate feelings, but not as much as groups posing symbolic threats.

8. Study 4

Study 4 aims to replicate Study's 3 findings using an alternative method. The similarities and differences in opinion around the social issues between participants and the political movement in Study's 3 manipulation, might lead participants to identify the party as either an in-group or an out-group respectively. To ensure that the effects revealed in Study 3 are not due to in-group/out-group identification, but

A. Aggressive tendencies



B. Hurting behaviors

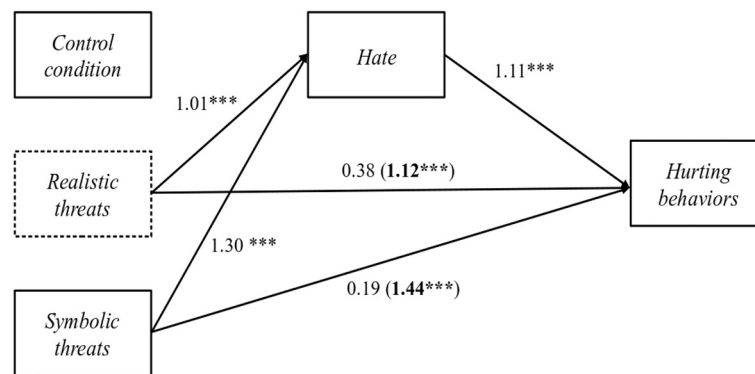


Fig. 4. Intergroup hate mediation effects of threat perceptions on aggressive tendencies (A), and hurting behaviors (B), Study 3.

Note. Dashed boxes represent the reference condition (i.e., realistic threats) of the $k - 1$ multi-categorical predictor. Significant relative indirect effects are in bold. $p < .05^*$, $p < .01^{**}$, $p < .001^{***}$.

explained by the different types of threat perceptions, we adapted a threat perception manipulation employed in previous threat research (e.g., Makashvili, Vardanashvili, & Javakhishvili, 2018; Rios, Sosa, & Osborn, 2018), to provide additional support to the previous findings while ruling out the possibility of this confound. In Study 4 we also include alternative measures of hate for examining the validity of the PHS.

8.1. Methods

8.1.1. Participants

Following the suggested sample size from the a-priori power analyses in Studies 2 and 3 (with the exact same parameters), we collected an initial online sample of $N = 400$ adults from the U.S. whose native language was English, and who did not participate previously in Studies 1, 2, or 3, from the Prolific Academic pool. All participants passed at least one attention check. Hurting behavior scores from six participants who failed the Tangram task check were coded as missing. After screening for multivariate outliers using Mahalanobis distances, we removed 27 meaningful outliers, obtaining a final sample of $N = 373$ adult participants, 35.1% males (age $M = 40.7$, $SD = 13.3$), 63.3% females (age $M = 39.9$, $SD = 15.3$), and 1.61% other gender (age $M = 28.7$, $SD = 8.73$), with different occupations and educational levels. Sensitivity power analyses conducted with our final sample using G*power (Faul et al., 2007), with $\alpha = 0.05$ and $1 - \beta = 0.80$, yielded a minimum detectable effect size of $f = 0.16$.

8.1.2. Procedure, design and analytical strategy

The procedure, design and analytical strategy of Study 4 were the

same as Studies 2 and 3, except for the manipulation (described below). The study took 13.15 min on average to be completed, and participants were paid £1.50 for their participation. Missing data were pairwise excluded from the analyses.

8.1.3. Manipulations and measures

To test the effect of the different types of threat perceptions we designed a fictional news article describing the arrival and settlement of a group of immigrants from a fictional country to the U.S. While the layout and story line were the same across conditions, we created three different versions associated with a) realistic threats, b) symbolic threats, and c) no threats. In the *realistic threats* condition, the immigrant group was depicted as posing an increase in job competition with locals, an increase in taxes covering the costs of their incorporation, and an increase in criminality, aiming to capture the breadth of different kinds of realistic threats. Similarly, in the *symbolic threats* condition, the immigrant group was depicted as endorsing cultural values divergent from the mainstream U.S. values regarding freedom of expression, choice, religion, and marriage. Finally, in the *no threat/control* condition, we included historical information about the group of immigrants (e.g., their language, origin, etc.) without making references to any kind of threat. The complete manipulated news articles are available in the online supplementary materials.

We employed the same measures from Study 3, but this time aimed towards the fictional immigrant group (and one hypothetical member in the aggression measures). Accordingly, participants answered the two manipulation checks first; then the hate scale ($\alpha = 0.93$); the aggressive tendencies measure (i.e., the voodoo doll task); and the hurting behaviors measure (i.e., the Tangram task). The aggression measures were

presented in a random and balanced order.

To examine the adequacy of the PHS for measuring hate, we included two additional hate measures to assess their convergent validity: The Triangular Hate Scale (THS) ($\alpha = 0.98$) (Sternberg & Sternberg, 2008), previously validated in US and German samples (Weis, 2006), and a Hate single item (in a scale from 1 = *not at all*, to 7 = *totally*, how much do you feel you hate the [immigrant group]?). The PHS showed a positive strong correlation with the THS, $r = 0.90$, $p < .001$, 95%CI [0.87, 0.91], and although the hate single item showed poor performance, it also showed to be positively correlated with the PHS, $r = 0.69$, $p < .001$, 95%CI [0.64, 0.74]. Thus, our PHS measure showed excellent convergent validity. Confirmatory Factor Analysis (CFA) testing the covariance between the PHS and the THS provide additional support, and all the results reported here replicate using the THS (see supplementary materials).

8.2. Results

8.2.1. Manipulation checks

In the realistic threats condition, participants perceived more threats to their goals ($M = 3.02$, $SD = 1.92$), than to their values ($M = 1.98$, $SD = 1.41$), $t(126) = 8.28$, $p < .001$, $d = 0.74$. In the symbolic threats condition, participants perceived more threats to their values ($M = 2.37$, $SD = 1.60$), than to their goals ($M = 2.02$, $SD = 1.30$), $t(117) = -2.81$, $p = .006$, $d = -0.26$. In the control condition participants perceived slightly more threats to their goals ($M = 1.97$, $SD = 1.37$), than to their values ($M = 1.79$, $SD = 1.32$), $t(127) = 2.84$, $p = .005$, $d = 0.25$. These results indicate that the manipulation was successful.

8.2.2. Effects of threat perceptions on hate

Results revealed a main effect of the threat manipulation on hate, $F(2, 370) = 12.72$, $p < .001$, $f = 0.26$. As predicted, post-hoc marginal means contrasts indicated that symbolic threats elicit higher levels of hate ($M = 2.34$, $SE = 0.07$) than realistic threats ($M = 1.99$, $SE = 0.07$), $B = -0.35$, $SE = 0.10$, $t = -3.51$, $p = .001$, $d = -0.45$, 95% CI [-0.58, -0.12]. Symbolic threats induced higher levels of hate than the control condition ($M = 1.85$, $SE = 0.07$), $B = -0.49$, $SE = 0.10$, $t = -4.92$, $p < .001$, $d = -0.63$, 95% CI [-0.72, -0.25]. Hate scores were not significantly different between the realistic threats and the control condition, $B = -0.12$, $SE = 0.09$, $t = -1.43$, $p = .329$, $d = -0.18$, 95% CI [-0.37, 0.09]. Results are summarized in Fig. 2G.

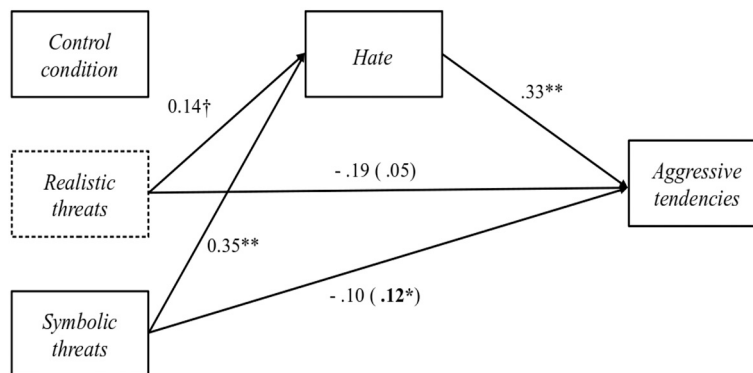
8.2.3. Associations between hate and aggression

Results showed an effect of hate on aggressive tendencies, $F(1, 371) = 8.12$, $p = .005$, $f = 0.15$, suggesting that hate significantly predicts aggressive tendencies, $B = 0.30$, $SE = 0.10$, $t = 2.85$, $p = .005$, 95% CI [0.09, 0.50]. Similarly, results revealed a significant effect of hate on hurting behaviors, $F(1, 365) = 175.4$, $p < .001$, $f = 0.35$, suggesting that hate significantly predicts hurting behaviors, $B = 0.51$, $SE = 0.08$, $t = 6.72$, $p < .001$, 95% CI [0.36, 0.66]. Results are summarized in Fig. 2H and I.

8.2.4. Mediation effects of hate between threat perceptions and aggression

Results indicate that as compared to realistic threats, symbolic threats significantly predict hate feelings, $B = 0.35$, $SE = 0.10$, $z = 3.42$, $p = .001$, 95% CI [0.16, 0.55], and with respect to the control condition, realistic threats does not predict hate, $B = 0.14$, $SE = 0.10$, $z = 1.41$, $p = .158$, 95% CI [-0.33, 0.06]. The aggressive tendencies model (Fig. 5A),

A. Aggressive tendencies



B. Hurting behaviors

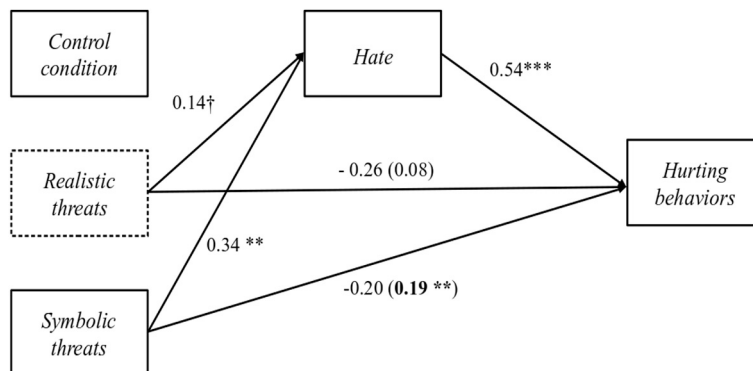


Fig. 5. Intergroup hate mediation effects of threat perceptions on aggressive tendencies (A), and hurting behaviors (B), Study 4.

Note. Dashed boxes represent the reference condition (i.e., realistic threats) of the $k - 1$ multi-categorical predictor. Significant relative indirect effects are in bold. $p < .05^*$, $p < .01^{**}$, $p < .001^{***}$.

revealed that hate significantly predict aggressive tendencies, $B = 0.33$, $SE = 0.12$, $z = 2.74$, $p = .006$, 95% CI [0.09, 0.57]. In line with our predictions, we found a non-significant relative direct effect of realistic threats on aggressive tendencies, $B = -0.19$, $SE = 0.23$, $z = 0.85$, $p = .397$, 95% CI [-0.21, 0.65], and a non-significant relative indirect effect of realistic threats on aggressive tendencies through hate, $B = 0.05$, $SE = 0.04$, $z = -1.30$, $p = .195$, 95% CI [-0.13, 0.01]. On the other hand, the relative direct effect between symbolic threats and aggressive tendencies was nonsignificant, $B = -0.10$, $SE = 0.19$, $z = -0.52$, $p = .599$, 95% CI [-0.49, 0.27]. As predicted, however, results showed a significant relative indirect effect of symbolic threats on aggressive tendencies through hate, $B = 0.12$, $SE = 0.06$, $z = 2.05$, $p = .04$, 95% CI [0.03, 0.26], suggesting that symbolic threats indirectly predict aggressive tendencies through intergroup hate.

The hurting behaviors model (Fig. 5B) revealed that hate also significantly predicts hurting behaviors, $B = 0.54$, $SE = 0.12$, $z = 4.72$, $p < .001$, 95% CI [0.34, 0.80]. Results revealed a non-significant relative direct effect of realistic threats on hurting behaviors, $B = 0.26$, $SE = 0.16$, $z = 0.16$, $p = .872$, 95% CI [-0.28, 0.36], and in line with our predictions a non-significant relative indirect effect of realistic threats on hurting behaviors through hate, $B = 0.08$, $SE = 0.05$, $z = 1.44$, $p = .150$, 95% CI [-0.02, 0.19]. Results also yielded a non-significant relative direct effect between symbolic threats and hurting behaviors, $B = -0.20$, $SE = 0.14$, $z = -1.41$, $p = .159$, 95% CI [-0.48, 0.08]. In line with predictions, we found a significant relative indirect effect of symbolic threats on hurting behaviors through hate, $B = 0.19$, $SE = 0.07$, $z = 2.69$, $p = .007$, 95% CI [0.08, 0.35].

8.3. Discussion

Study 4 provides support to our main predictions employing an alternative and previously used outgroup threat manipulation. Once again, compared to realistic threats, symbolic threats predict intergroup feelings of hate more strongly, and hate shows positive significant associations with aggressive tendencies and hurting behaviors. Additionally, as expected, intergroup hate mediates the associations between symbolic (but not realistic) threats and the two aggressive outcomes, mirroring Study's 2 findings. This provides additional support to the idea that the significant difference in hate ratings between the realistic and the control conditions in Study 3 (not evidenced in Studies 2 and 4), might explain the unexpected significant mediation of hate between realistic threats and aggression in Study 3.

9. General discussion

Consistent with our primary hypothesis, Study 1 revealed that compared to realistic threats, symbolic threats predict interpersonal and intergroup hate more strongly. As expected, results also revealed that hate significantly predicts aggression across targets, in particular aggressive tendencies and hurting behaviors. Study 2 revealed that as compared to realistic threats, symbolic threats elicit higher levels of interpersonal hate, which in turn mediates the relationships between symbolic (but not realistic) threats and aggressive outcomes. Study 3 revealed that participants in the symbolic threats condition report higher levels of intergroup hate than participants in the realistic threats condition. Somewhat unexpectedly, however, results showed that both realistic and symbolic threats predict intergroup hate, and intergroup hate positively mediates the relations between both symbolic and realistic threat perceptions and aggressive outcomes. Finally, Study 4 revealed that symbolic threats predict intergroup hate to a higher degree than realistic threats, that hate predicts aggressive tendencies and hurting behaviors, and as predicted, that hate mediates the association between symbolic (but not realistic) threats and aggression. Taken together, these findings suggest three main conclusions. First, compared to realistic threats, symbolic threats more strongly predict feelings of hate. Second, hate predicts overt forms of aggression, but may not

(consistently) predict more subtle forms of aggression such as social hostility. And third, although symbolic threats consistently predict both interpersonal and intergroup hate, realistic threats may also contribute to hate and aggression between groups, especially when outgroups pose concrete realistic threats to individuals.

9.1. Implications for hate and threat research

The present research poses at least three broad implications. First, our findings contribute to a more fine-grained conceptualization of hate by providing evidence that symbolic threats predict hate, which is associated with aggression. Second, by uncovering the relevance of symbolic threats for hate, our findings illuminate a hate's potential function of protecting individuals from targets posing threats to their core values and identity, even to a greater extent than threats to their resources or safety. As such, the present findings bring us one step closer towards mapping the psychological processes associated with hate. Third, our findings make a novel contribution to the ITT and the threat-based functional models of prejudice (Stephan et al., 2009). Apart from some evidence showing that realistic threats elicit anger while symbolic threats elicit anger and disgust towards out-groups (Matthews & Levin, 2012), neither hate or the co-occurrence of hate component emotions (Matsumoto et al., 2016, 2017; Sternberg & Sternberg, 2008; Sternberg, 2003) have been previously addressed within this framework (Aubé & Ric, 2019; Cottrell & Neuberg, 2005; Johnston & Glasford, 2014).

Adopting such a co-occurrence approach to hate pose some advantages. First, discrete emotions are rarely activated in isolation and people often experience a mix of emotions (Lange, Dalege, Borsboom, van Kleef, & Fischer, 2020), especially in complex phenomena such as prejudice. And second, a co-occurrence approach to hate brings versatility in the analyses of particular circumstances under which one component emotion or blend of emotions is more salient. For example, in our supplementary analyses where we analyze hate's component emotions separately, we found that in combination with symbolic threats, realistic threats also evoke anger at the interpersonal level and disgust at the intergroup level. In line with Sternberg's (2003) typology, where seven types of hate are derived from all the possible combination of the three component emotions, this approach can offer a more nuanced picture of the symbolic threat – hate – aggression pathway which might help us understand the origins and persistence of prejudice and conflict (Staub, 2005).

Future research may more closely explore the mechanisms underlying the effect of symbolic threats on hate. For instance, the strong emotionality with which people endorse their values, worldviews, and moral convictions (Skitka, Bauman, & Sargis, 2005), may explain the proportional strong emotional reactions to symbolic threat, facilitating hate. Similarly, it is possible that symbolic threats constitute threats to one's moral identity (cf., Aquino & Reed, 2002), leading to strong emotional reactions to protect one's self-concept and preserve internal coherence. Furthermore, symbolic threats may be experienced as stable in time on the basis of negative and non-malleable dispositional attributions towards the targets (e.g., Baumeister, 1997; Hutcherson & Gross, 2011), eliciting proportional enduring hate feelings with the prospective function of keeping individuals prepared to face the threats in the long run (Roseman & Steele, 2018).

Finally, future research may further explore the effect of realistic threats on intergroup hate. Supplementary mediation analyses using the threat perceptions scale as predictors in Study 1 (see Fig. S4) showed that realistic threats do not significantly predict intergroup hate in line with our hypothesis, Study's 2 and 4 results, and contrary to the mediation results in Study 3. This inconsistency may be explained by the significant contrast between the realistic threats condition and the control condition without threats in Study 3, compared to the only measured realistic threat perceptions in the supplementary analyses. Nevertheless, Study's 3 findings suggest that it would be premature to exclude a possible role of realistic threats in intergroup hate. Groups are

not only more powerful in number, but also in terms of decision making, establishing the status quo, or translating ideologies and values into concrete circumstances that pose direct realistic threats to individuals. It is also possible that severe realistic threats posed by out-groups (e.g., war, invasion), might give rise to hate feelings. More concretely, we regard the roles of out-group power, status, and size in shaping intergroup hate a promising avenue of future research.

9.2. Strengths and limitations

The present research has some strengths. First, both correlational and experimental evidence support the current propositions, formalizing previous theorizing about hate (Sternberg, 2003). Second, we provide conceptual and statistical ground to the hate construct following a psychometric model of emotions (Lange et al., 2020), and we conduct convergent validity analyses of different hate measures. Third, unlike the vast majority of the threat-based research measuring threat only as an observed variable with the threat perceptions scale (Stephan & Stephan, 2017), we manipulated threat perceptions in Studies 2, 3 and 4, providing experimental evidence of the effects of realistic and symbolic threats on hate. Indeed, meta-analytical reviews point to the remarkably scarce experimental research reported in the ITT literature (Riek, Mania, & Gaertner, 2006; Rios et al., 2018). Finally, instead of self-report our study included behavioral measures of aggression using lab-based paradigms (McCarthy & Elson, 2018). In this way, we obtained relatively valid and robust estimates of aggression.

The present research also has some limitations. First, we only addressed the two broad categories of realistic and symbolic threats as antecedents of hate and there is still room for further specification. For instance, different kinds of symbolic threats (e.g., to moral values, worldviews, identity, honor, political ideology, religious beliefs, or beliefs about justice) may pose differential effects for the development of hate. Similarly, there is a wide array of realistic threats ranging from small economic costs to life threatening situations, posing further existential threats (cf., Halperin et al., 2009; Hirschberger, Ein-Dor, Leidner, & Saguy, 2016; Wohl, Branscombe, & Reysen, 2010). Second, although we strived to operationalize and manipulate threat perceptions in the most possible controlled ways, the conceptual boundaries between symbolic and realistic threats may be blurred under some particular circumstances, and future research should consider this for reaching more refined distinctions. Finally, while Studies 2 to 4 revealed a causal effect of threat perceptions on hate feelings, the link between hate and aggression was examined and supported by correlations across all four studies. Future research therefore may gather more definitive evidence for our mediation model by experimentally manipulating hate, and testing its effects on aggressive behaviors. Further specification of hate's antecedents, components, and aggressive outcomes, would be of immense value for evaluating better predictive models and preventing hate cycles.

9.3. Concluding remarks

Our findings constitute an initial step towards examining whether hate may be a result of symbolic and realistic threats. The key finding across studies is that hate is better predicted by symbolic than realistic threats, suggesting that hate is triggered by conflicting interests but especially by threats to people's values and identity. Our findings also provide evidence for the association between hate and aggressive behaviors intending to block target's goals and cause harm. In this way, the model we propose may illuminate the dynamics of the threat-hate-aggression link. The present findings also point at psychological differences between interpersonal and intergroup hate in terms of the reactivity to realistic and symbolic threats, suggesting that realistic threats may also play a role in eliciting intergroup hate. These insights might serve as a springboard towards an empirical understanding of the antecedents and consequences of hate, a feeling that often seems to be

deeply rooted in strong and persistent differences between own and target's core moral values and identity – differences that are key for understanding both interpersonal and intergroup conflict.

Open practices

All materials, data, analyses, and preregistrations are available on the open science framework (<https://osf.io/ksx74>).

Funding

This work was supported by the Administrative Department of Science, Technology and Innovation (Colciencias) [Scholarship number 783, 2017].

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jesp.2022.104393>.

References

- Aquino, K., & Reed, A. I. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, 83(6), 1423–1440. <https://doi.org/10.1037/0022-3514.83.6.1423>
- Aubé, B., & Ric, F. (2019). The sociofunctional model of prejudice: Questioning the role of emotions in the threat-behavior link. *International Review of Social Psychology*, 32(1), 1–15. <https://doi.org/10.5334/irsp.169>
- Aumer, K., Bahn, A. C. K., Janicki, C., Guzman, N., Pierson, N., Strand, S. E., & Totlund, H. (2016). Can't let it go: Hate in interpersonal relationships. *Journal of Relationships Research*, 7, Article e2. <https://doi.org/10.1017/jrr.2016.2>
- Baumeister, R. (1997). *Evil: Inside human cruelty and violence*. W.H. Freeman.
- Baumeister, R., & Butz, D. (2005). Roots of hate, violence, and evil. In R. Sternberg (Ed.), *The psychology of hate* (pp. 87–102). American Psychological Association.
- Beck, A. T. (2000). Prisoners of hate: The cognitive basis of anger, hostility, and violence. *Perennial*.
- Ben-Ze'ev, A. (2000). *The subtlety of emotions*. MIT Press.
- Chetty, N., & Alathur, S. (2018). Hate speech review in the context of online social networks. *Aggression and Violent Behavior*, 40, 108–118. <https://doi.org/10.1016/j.avb.2018.05.003>
- 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), 770–789. <https://doi.org/10.1037/0022-3514.88.5.770>
- Cowling, M., Anderson, J., & Ferguson, R. (2019). Prejudice-relevant correlates of attitudes towards refugees: A meta-analysis. *Journal of Refugee Studies*, 32(3), 502–524. <https://doi.org/10.1093/jrs/fev062>
- Dewall, C. N., Finkel, E. J., Lambert, N. M., Slotter, E. B., Bodenhausen, G. V., Pond, R. S., ... Fincham, F. D. (2013). The voodoo doll task: Introducing and validating a novel method for studying aggressive inclinations. *Aggressive Behavior*, 39(6), 419–439. <https://doi.org/10.1002/ab.21496>
- Elnakouri, A., Hubley, C., & McGregor, I. (2022). Hate and meaning in life: How collective, but not personal, hate quells threat and spurs meaning in life. *Journal of Experimental Social Psychology*, 98, Article 104227. <https://doi.org/10.1016/j.jesp.2021.104227>
- Esses, V., & Hamilton, L. (2021). Xenophobia and anti-immigrant attitudes in the time of COVID-19. *Group Processes and Intergroup Relations*, 24(2), 253–259. <https://doi.org/10.1177/1368430220983470>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Fischer, A., & Giner-Sorolla, R. (2016). Contempt: Derogating others while keeping calm. *Emotion Review*, 8(4), 346–357. <https://doi.org/10.1177/1754073915610439>
- Fischer, A., Halperin, E., Canetti, D., & Jasini, A. (2018). Why we hate. *Emotion Review*, 175407391775122. <https://doi.org/10.1177/1754073917751229>
- Fromm, E. (1992). *The anatomy of human destructiveness*. H. Holt.
- Glaser, J., Dixit, J., & Green, D. P. (2002). Studying hate crime with the internet: What makes racists advocate racial violence? *Journal of Social Issues*, 58(1), 177–193. <https://doi.org/10.1111/1540-4560.00255>
- Halperin, E., Canetti-Nisim, D., & Hirsch-Hoefler, S. (2009). The central role of group-based hatred as an emotional antecedent of political intolerance: Evidence from Israel. *Political Psychology*, 30(1), 93–123. <https://doi.org/10.1111/j.1467-9221.2008.00682.x>
- Halperin, E. (2008). Group-based hatred in intractable conflict in Israel. *The Journal of Conflict Resolution*, 52(5), 713–736. <https://doi.org/10.1177/0022002708314665>
- Hayes, A. F., & Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology*, 67(3), 451–470. <https://doi.org/10.1111/bmsp.12028>
- Hirschberger, G., Ein-Dor, T., Leidner, B., & Saguy, T. (2016). How is existential threat related to intergroup conflict? Introducing the multidimensional existential threat

- (MET) model. *Frontiers in Psychology*, 7(DEC). <https://doi.org/10.3389/fpsyg.2016.01877>
- Hjerm, M., & Nagayoshi, K. (2011). The composition of the minority population as a threat: Can real economic and cultural threats explain xenophobia? *International Sociology*, 26(6), 815–843. <https://doi.org/10.1177/0268580910394004>
- Hutcherson, C. A., & Gross, J. J. (2011). The moral emotions: A social-functional account of anger, disgust, and contempt. *Journal of Personality and Social Psychology*, 100(4), 719–737. <https://doi.org/10.1037/a0022408>
- Johnston, B. M., & Glasford, D. E. (2014). A threat-emotion profile approach to explaining active versus passive harm in intergroup relations. *Social Psychology*, 45(5), 399–407. <https://doi.org/10.1027/1864-9335/a000199>
- Kucuk, S. U. (2016). *What is hate? In brand hate* (pp. 1–15). Springer International Publishing. https://doi.org/10.1007/978-3-319-41519-2_1
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13). <https://doi.org/10.18637/jss.v082.i13>
- Lange, J., Dalege, J., Borsboom, D., van Kleef, G. A., & Fischer, A. H. (2020). Toward an integrative psychometric model of emotions. *Perspectives on Psychological Science*, 174569161989505. <https://doi.org/10.1177/1745691619895057>
- Lenth, R. (2020). emmeans: Estimated marginal means, aka least-squares means. (R package version 1.5.3.). <https://cran.r-project.org/package=emmeans>
- Makashvili, A., Vardanashvili, I., & Javakhishvili, N. (2018). Testing intergroup threat theory: Realistic and symbolic threats, religiosity and gender as predictors of prejudice. *Europe's Journal of Psychology*, 14(2), 464–484. <https://doi.org/10.5964/EJOP.V14I2.1483>
- Martínez, C. A., van Prooijen, J.-W., & Van Lange, P. A. M. (2022). Hate: Toward understanding its distinctive features across interpersonal and intergroup targets. *Emotion*, 22(1), 46–63. <https://doi.org/10.1037/emo0001056>
- Matsumoto, D., Hwang, H. C., & Frank, M. G. (2016). The effects of incidental anger, contempt, and disgust on hostile language and implicit behaviors. *Journal of Applied Social Psychology*, 46(8), 437–452. <https://doi.org/10.1111/jasp.12374>
- Matsumoto, D., Hwang, H. C., & Frank, M. G. (2017). Emotion and aggressive intergroup cognitions: The ANCODI hypothesis. *Aggressive Behavior*, 43(1), 93–107. <https://doi.org/10.1002/ab.21666>
- Matthews, M., & Levin, S. (2012). Testing a dual process model of prejudice: Assessment of group threat perceptions and emotions. *Motivation and Emotion*, 36(4), 564–574. <https://doi.org/10.1007/s11031-012-9280-y>
- McCarthy, R. J., & Elson, M. (2018). A conceptual review of lab-based aggression paradigms. *Collabra: Psychology*, 4(1), 4. <https://doi.org/10.1525/collabra.104>
- Opatow, S., & McClelland, S. I. (2007). The intensification of hating: A theory. *Social Justice Research*, 20(1), 68–97. <https://doi.org/10.1007/s11211-007-0033-0>
- Riek, B. M., Mania, E. W., & Gaertner, S. L. (2006). Intergroup threat and outgroup attitudes: A Meta-analytic review. *Personality and Social Psychology Review*, 10(4), 336–353. https://doi.org/10.1207/s15327957pspr1004_4
- Rios, K., Sosa, N., & Osborn, H. (2018). An experimental approach to intergroup threat theory: Manipulations, moderators, and consequences of realistic vs. symbolic threat. *European Review of Social Psychology*, 29(1), 212–255. <https://doi.org/10.1080/10463283.2018.1537049>
- Roseman, I. J., & Steele, A. K. (2018). Concluding commentary: Schadenfreude, Gluckschmerz, jealousy, and hate—What (and when, and why) are the emotions? *Emotion Review*, 10(4), 327–340. <https://doi.org/10.1177/1754073918798089>
- Rossee, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(1), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Saleem, M., Anderson, C. A., & Barlett, C. P. (2015). Assessing helping and hurting behaviors through the tangram help/hurt task. *Personality and Social Psychology Bulletin*, 41(10), 1345–1362. <https://doi.org/10.1177/0146167215594348>
- Shepherd, L., Fasoli, F., Pereira, A., & Branscombe, N. (2018). The role of threat, emotions, and prejudice in promoting collective action against immigrant groups. *European Journal of Social Psychology*, 48(4), 447–459. <https://doi.org/10.1002/EJSP.2346>
- Silva, R., Oliveira, J., Dias, C., Pinto, I., & Marques, J. (2018). How inclusive policies shape prejudice versus acceptance of refugees: A Portuguese study. *Peace and Conflict*, 24(3), 296–305. <https://doi.org/10.1037/PAC0000314>
- Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005). Moral conviction: Another contributor to attitude strength or something more? *Journal of Personality and Social Psychology*, 88(6), 895–917. <https://doi.org/10.1037/0022-3514.88.6.895>
- Staub, E. (2005). The origins and evolution of hate, with notes on prevention. In R. Sternberg (Ed.), *The psychology of hate* (pp. 51–66). American Psychological Association. <https://doi.org/10.1037/10930-003>
- Staub, E. (2011). *Overcoming evil: Genocide, violent conflict, and terrorism*. Oxford University Press.
- Stephan, W., & Stephan, C. (2000). An integrated threat theory of prejudice. In S. Oskamp (Ed.), *Reducing prejudice and discrimination* (pp. 23–45). Lawrence Erlbaum Associates. <https://doi.org/10.4324/9781410605634-7>
- Stephan, W., & Stephan, C. (2017). Intergroup threat theory. In *The international encyclopedia of intercultural communication* (pp. 1–12). Inc: John Wiley & Sons. <https://doi.org/10.1002/9781118783665.ieicc0162>
- Stephan, W., Ybarra, O., & Morrison, K. (2009). Intergroup threat theory. In T. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 43–60). Psychology Press.
- Sternberg, S. (2008). *The nature of hate*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511818707>
- Sternberg, S. (2003). A duplex theory of hate: Development and application to terrorism, massacres, and genocide. *Review of General Psychology*, 7(3), 299–328. <https://doi.org/10.1037/1089-2680.7.3.299>
- Van Bavel, J., Ray, J., & Cunningham, W. (2018). The psychology of hate: Moral concerns differentiate hate from dislike. *PsyArXiv*. <https://doi.org/10.31234/OSF.IO/X9Y2P>
- Van Doesum, N., Van Prooijen, J.-W., Verburgh, L., & Van Lange, P. A. M. (2016). Social hostility in soccer and beyond. *PLoS ONE*, 11(4). <https://doi.org/10.1371/journal.pone.0153577>
- Van Doorn, J. (2018). Anger, feelings of revenge, and hate. *Emotion Review*, 175407391878326. <https://doi.org/10.1177/1754073918783260>
- Van Lange, P. A. M., & Van Doesum, N. (2015). Social mindfulness and social hostility. *Current Opinion in Behavioral Sciences*, 3, 18–24. <https://doi.org/10.1016/j.cobeha.2014.12.009>
- Weis, K. (2006). *Explorations of the duplex theory of hate*. Berlin: Logos Verlag.
- Wirtz, C., Van Der Pligt, J., & Doosje, B. (2016). Negative attitudes toward muslims in the Netherlands: The role of symbolic threat, stereotypes, and moral emotions. *Peace and Conflict*, 22(1), 75–83. <https://doi.org/10.1037/pac0000126>
- Wohl, M. J. A., Branscombe, N. R., & Reysen, S. (2010). Perceiving your Group's future to be in jeopardy: Extinction threat induces collective angst and the desire to strengthen the Ingroup. *Personality and Social Psychology Bulletin*, 36(7), 898–910. <https://doi.org/10.1177/0146167210372505>
- Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, Bias, and solution propriety. *Educational and Psychological Measurement*, 73(6), 913–934. <https://doi.org/10.1177/0013164413495237>
- Zeki, S., & Romaya, J. P. (2008). Neural correlates of hate. *PLoS One*, 3(10), Article e3556. <https://doi.org/10.1371/journal.pone.0003556>