1-1-2022

The Psychology of Hate: Moral Concerns Differentiate Hate from Dislike

Clara Pretus
Universitat Autònoma de Barcelona

Jennifer L. Ray
Takeda Pharmaceuticals U.S.A., Inc.

Yael Granot
Loyola University of Chicago, ygranot@smith.edu

William A. Cunningham
University of Toronto

Jay J. Van Bavel
New York University

The psychology of hate:

Moral concerns differentiate hate from dislike

Clara Helena Pretus Gomez
Universitat Autonoma de Barcelona

Jennifer L. Ray
MindGym

Yael Granot
Loyola University

William A. Cunningham
University of Toronto

Jay J. Van Bavel
New York University

Acknowledgements: The authors thank Stacey Fox, Doreen Kinley, Bryan Moeller, Erika Price, and Jessica Wilkins for assisting with data collection and coding. We also wish to thank members of the NYU Social Perception and Evaluation Lab, UT Computational Affective Neuroscience Lab, and OSU Group for Attitudes and Persuasion for constructive comments. This research was partially supported by a SSHRC Canada Graduate Scholarship and NSF Award (#1349089) to Jay Van Bavel, a NSF Graduate Research Fellowship to Yael Granot, and SSHRC Standard Research Grant and Mershon Center for International Security Studies Grant to William Cunningham. Parts of this research were presented by Jennifer Ray to the Association for Psychological Science in 2014, to the Attitudes Preconference at the Society for Personality and Social Psychology in 2014, and to the Society for Personality and Social Psychology in 2015. Last authors WAC and JVB contributed equally to this paper and last authorship was determined by the luck of the draw.

Responsibilities: JVB and WAC designed S1-3; JVB, JLR, and YG designed S4; JVB programmed and conducted S1-3; YG and JLR collected data and analyzed S4; JVB, JLR, and YG analyzed S1-3; JVB, JLR and YG drafted the manuscript; JVB, HE, JM, CG and WAC edited the manuscript.

Contact information: Correspondence concerning this article should be addressed to Jay J. Van Bavel, New York University, Psychology Department, 6 Washington Pl, New York, NY 10003, email: jay.vanbavel@nyu.edu

Abstract

We investigated whether any differences in the psychological conceptualization of hate and dislike were simply a matter of degree of *negativity* (i.e., hate falls on the end of the continuum of dislike) or also *morality* (i.e., hate is imbued with distinct moral components that distinguish it from dislike). In three lab studies in Canada and the US, participants reported disliked and hated attitude objects and rated each on dimensions including valence, attitude strength, morality, and emotional content. Quantitative and qualitative measures revealed that hated attitude objects were more negative than disliked attitude objects and associated with moral beliefs and emotions, even after adjusting for differences in negativity. In study four, we analyzed the rhetoric on real hate sites and complaint forums and found that the language used on prominent hate websites contained more words related to morality, but not negativity, relative to complaint forums. (137/150 words)

Keywords: hate, dislike, attitudes, morality, emotions

The psychology of hate:

Moral concerns differentiate hate from dislike

The underpinnings of the psychological state of hate have become increasingly relevant as hate-based crimes are on the rise internationally. The UK documented a 30% increase in hate crimes in 2016 (de Freytas-Tamura, 2017), and reported incidents of hate crimes in the United States similarly increased in 2016 compared to the previous year (Berman, 2017). Whereas hate as a legal concept counts on a necessarily specific definition ("intense or extreme dislike, aversion, loathing, antipathy, enmity or hostility against members of groups or classes of persons identified by protected characteristics," as per Brown, 2017), what people perceive and experience as "hate" at a psychological level is still not well understood. Legal and psychological concepts can be highly interconnected. For instance, legal definitions and procedures around sexual harassment and domestic violence have greatly evolved in the last decades thanks to a greater understanding of the psychological aspects involved in these crimes (Fitzgerald et al., 1997; Meier, 1993). Thus, advancing the understanding of the psychology of hate can provide valuable insights into how it can be best operationalized in the legal system. In the present work we were interested in examining the psychological correlates of our shared cultural meaning of "hate".

The concept of hate

Hate has been generally conceptualized as a negative emotional attitude (Allport et al., 1954; White, 1996) toward persons or groups who are considered to possess fundamentally negative traits (Allport et al., 1954; Ben-Ze'ev, 2000; Royzman et al., 2006; Sternberg, 2003). Similarly to other affective states (Frijda et al., 1991), hatred can be experienced both as an emotion ("acute hate") and as a sentiment or emotional attitude ("chronic hate") (Bartlett, 2005;

Halperin et al., 2012). Hate as an emotion has been described as an acute reaction to a significant event characterized by intense feelings, physical symptoms, and desire for immediate action (Sternberg, 2003, 2005). In this respect, many authors have attempted to define hate in contrast to anger, both in terms of appraisals and motivational goals (Allport et al., 1954; Royzman et al., 2006; Sternberg, 2005). Hated targets are viewed as innately evil and incapable of change, while anger focuses on the actions attributed to an agent (Ortony et al., 1990), who is perceived as malleable (Fischer & Roseman, 2007; Halperin, 2011). Thus, hated targets pose an existential threat to subjects (Ben-Ze'ev, 2000), who experience low levels of control (Fitness & Fletcher, 1993) and powerlessness (Fischer et al., 2018). As a result, haters seek to avoid, hurt, or eliminate the object of hate (Allport et al., 1954; Ben-Ze'ev, 2000; Martínez et al., 2022; Sternberg, 2005), while anger entails motivational goals related to the subject's desire to change the target (Fischer et al., 2018).

When hate is defined as an emotional attitude (versus hate as an emotion), it is a more stable disposition towards a hated object that relies significantly on cognition (Halperin et al., 2012). In this respect, hatred can be conceptually compared to dislike, which can be defined as a preference or negative disposition towards an object that influences our behavior (de Houwer & Hughes, 2020). Hatred and dislike have been historically conceptualized as two related concepts. For instance, German poet Johann Wolfgang von Goethe famously speculated that “Hatred is active, and envy passive dislike; there is but one step from envy to hate” (see Edwards, 1908) and Darwin (1872) argued that “Dislike easily rises into hatred”, suggesting that the two belong on a common conceptual spectrum.

From a modern psychological perspective, what we label dislike and hatred both share some characteristics in terms of their dispositional nature and negative valence. However, the things that we say that we hate, as opposed to dislike, are transmissible, lead to false attributions, and motivate violent crimes (Sternberg, 2005). We believe these differences may be understood by taking into account the moral dimension of hate.

Hate versus dislike: Does morality matter?

Preferences can transform into values through the process of moralization (Rozin, 1999). As a result, moral values, as compared to preferences, are more central to the self and have a greater ability to impact an individual's life and society at large in powerful ways (see also Skitka, 2002). In line with this, parent-to-child transmission has been found to be more consistent for moral values than preferences (Cavalli-Sforza et al., 1982). Framing an action as moral leads to more extreme judgements (van Bavel et al., 2012) and fosters prosocial behavior regardless of preferences (Capraro & Rand, 2017). The willingness to make extreme sacrifices is greater for highly moralized values (i.e., sacred values) compared to preferences (Atran et al., 2014). Thus, the distinctive status of hate as "the most destructive affective phenomenon in the history of human nature" (Royzman et al., 2006) could well emanate from its moralized nature.

There are different ways in which hatred could relate to morality. Some scholars suggest that hatred as an emotional disposition is linked to negative appraisals evoked by moral transgressions. For instance, Allport (1954) conceived of hate as a strong form of dislike that escalates in negativity but speculated that hatred cannot exist "unless something one values has been violated" (pg. 364). Along the same lines, others have argued that hatred is characterized as the negative evaluation of a target that is linked with a moral judgment (Royzman et al., 2006), that hate is rooted in seeing the hated target as morally deficient or as violating moral norms (Staub, 2004) and that hatred is a direct reaction to protracted harm perceived as deliberate, unjust, and stemming from an inner evil character of the hated individual or group (Halperin, 2008). These researchers argue that hate is inextricably linked to morality through negative moral appraisals.

Similarly, the Duplex Theory of Hate argues that hate as an emotion may be composed of other more basic moral emotions—contempt, anger, and disgust— which are triggered by moral transgressions (Sternberg, 2003). Particularly, transgressions to communal codes, including hierarchy, have been associated with contempt, transgressions to personal autonomy or individual rights have been associated with anger, and violations to purity/sanctity have been linked to disgust (see also Rozin, 1999). Therefore, there seems to be some consensus over the notion that hatred is elicited by moral transgressions, either through moral negative appraisals or through moral emotions. Despite the numerous accounts of hate as a moral sentiment, no studies to our knowledge have empirically tested whether hate is different from negative preferences (i.e., dislike) in the moral domain.

Empirical evidence to date

Despite the rich history of theorizing about the nature of hate—and the serious real-world implications of the topic—surprisingly little empirical research has investigated the psychological conceptualization of hate. One study found that hate is the sixth most frequently listed exemplar of the concept of emotion (behind happiness, anger, sadness, love, and fear), and therefore, among the most psychologically accessible emotional concepts to people (Fehr & Russell, 1984). Other work exploring hate in interpersonal contexts has found that people write significantly longer, detailed accounts of their experiences of hate, as compared to experiences of anger, jealousy, or love. However, when confronted with hypothetical emotion scenarios, participants were least accurate at identifying hate (Fitness & Fletcher, 1993). Thus, the limited literature on the psychology of hate underscores the everyday importance of the topic, as well as the ways in which hate is little understood by scientists and laypersons alike.

In attempting to understand the psychological experience of hate, researchers have investigated what other experiences or emotions are associated with hate. For example, participants who wrote about targets of hate often described friends, family, and acquaintances (56%), but rarely strangers (4%) (Aumer-Ryan & Hatfield, 2007), suggesting a link between hate and familiarity. More recent work has found that higher emotional arousal, personal threat perceptions, and attack-oriented behaviors are hate's distinctive features (see Martínez et al., 2022). Further, people who recalled an experience of hate for their partner reported feeling significantly less in control than when experiencing other negative emotions like anger (Fitness & Fletcher, 1993). Finally, experiencing hate may correspond with aversive health consequences, including increased blood pressure and immune system suppression (Dozier, 2002). Consequently, hate's inherent link to violence has prompted some researchers to begin studying hate as a public health concern (Krug, 2002).

Concerning the moral dimension of hate, research on hate groups suggests that members were particularly likely to cite symbolic, value-relevant issues as sources for their hate. Moreover, hate groups were also more likely to advocate violence toward the hated group in the face of threats of a moral nature in phone interviews (Green et al., 1999) and online chat rooms (Glaser et al., 2002). A recent set of studies found that moral values oriented around group preservation are predictive of the county-level prevalence of hate groups and associated with the belief that extreme behavioral expressions of prejudice against marginalized groups are justified (Hoover et al., 2021). As such, real world hate groups appear to be motivated by perceived moral transgressions, even if no actual transgression happened, and this might foster real-world aggression against vulnerable groups.

Overview

The current research examined the role of morality in the psychology of hate. We conducted four studies designed to investigate whether the difference between how people conceptualize hate and dislike is simply a matter of *intensity* (i.e., hate is merely more negative than dislike) or also morality (i.e., hate is imbued with additional psychological ingredients in the moral domain). Do people differentiate hate from dislike strictly in terms of intensity, such that dislike is negative and hate is *extremely* negative on the same continuum? Or do people differentiate hate from dislike along the moral dimension above and beyond negativity? Most theories focus on differences in negativity or morality without directly testing them concurrently. As such, it is possible that both accounts are correct to some degree.

To address this question, we tested two primary hypotheses—which we termed the *intensity hypothesis* and the *morality hypothesis*. In line with theorizing by Allport (1954) and Ben Ze'ev (2000), hate may best be conceptualized by people as a negative evaluation. For this reason, we hypothesized that if the difference between psychological conceptualizations of hate and dislike is largely a matter of intensity, then hated attitude objects would be rated as more negative than disliked attitude objects (*intensity hypothesis*). In contrast, following the many theorists who claimed that hate is connected to morality (Allport et al., 1954; Royzman et al., 2006; Staub, 2004), hate may differ in kind from dislike along a moral dimension (*morality hypothesis*). We hypothesized that hated attitude objects would be associated with more moral emotions, such as anger, contempt, and disgust (Rozin & Fallon, 1987) and be rated as more connected to core moral beliefs than disliked attitude objects. These hypotheses do not need to be competitive, in fact, we predicted that hated attitude objects would be both more negative and moral than disliked objects.

Across four studies we tested these hypotheses using data from controlled laboratory experiments as well as examined the language used online by prominent hate groups. Of note, our lab samples include Canadian and US students. Therefore, the conceptualizations of hate, dislike, and morality discussed in this work are situated within a North American cultural context. Different cultures may have different environments, social structures, or mental states associated with morality, intent, and responsibility that may shape these concepts (see Henrich, 2022). In the lab, we asked participants to generate their own hated and disliked attitude objects and rate them on dimensions of valence, emotional content, motivation, and morality (Studies 1-3). We also asked participants to self-reflect on how they understood the difference between their relationship to their hated and disliked attitude objects and we coded their qualitative responses (Study 1). The research built upon previous studies of the phenomenology of hate (Aumer-Ryan & Hatfield, 2007) by including established measures of morality and attitude strength (Skitka et al., 2005) to conduct quantitative tests of our key hypotheses.

It is important to note that our research largely focuses on lay theories of hate. Thus, any conclusions are limited to the shared semantic content of what people call “hate”. For instance, previous research found that 31% of participants consider hate to be roughly synonymous with extreme dislike (Aumer-Ryan & Hatfield, 2007). In this respect, whereas most theoretical accounts conceptualize hate as an emotional disposition towards persons or groups, people often use the expression “hate” to refer to objects or concepts. While we did not force a specific meaning upon participants regarding “hate”, we separately asked about hate in relation to people/groups versus concepts/beliefs in Study 3. From our perspective, this lack of constraint in meaning makes any formal test of qualitative differences—over-and-above general dislike—more conservative. However, it also introduces a constraint on our ability to generalize the results to more ecologically valid contexts. As such, we addressed this limitation in our final study by analyzing the rhetoric on real hate group websites (vs. complaint websites).

To support the ecological validity of our laboratory research, we conducted a content analysis of websites run by real hate groups, comparing the linguistic content of hate group websites with online consumer complaint forums and employee complaint forums (Study 4). Hate group websites were selected to sample expressions of hate, while complaint forums were selected to sample expressions of dislike in relation to either products (consumer complaint forums) or people/corporations (employee complaint forums). Although hate groups are publicly recognized as such, many of these groups adamantly fight against the hate group label and feel they are merely expressing their own deeply held values (an issue we discuss in depth below). Thus, the content of these websites is more likely to embody the natural occurrence of hateful rhetoric and values in the real world. This also expands our research to the domain of intergroup relations since hate groups are defined as organizations that—based on their official statements of principles, the statements of their leaders, or their activities – hold beliefs or engage in practices that attack or malign entire classes of people, typically for their immutable characteristics (SPLC, 2017). Together, these studies served as an initial exploration into how people evaluate and represent their own hatred.

Study 1: Hate versus Dislike

Our first study sought to investigate whether differences in people's conceptualization of hate and dislike are simply a matter of *intensity* (i.e., hate is more negative than dislike) or also *morality* (i.e., hate involves additional moral content). To address this question, we explored what types of attitude objects were associated with hate as compared to dislike, and whether those objects differed as to the valence, emotional content, and motivational and moral concerns they promoted. We used analyses of both quantitative ratings and qualitative reports to compare hate vs. dislike on these critical dimensions.

Method

The materials, data and analysis code are available on the Open Science Framework website (osf.io/5u6my). A power analysis run with the R package 'simr' (Green et al., 2016) showed that, based on 1000 simulations and an alpha = 0.05, recruiting 170 participants would enable the detection of small effects (e.g. 0.1) for the interaction between within-subject attitude object type (hated vs. disliked object) and order (hate first vs. dislike first) with a statistical power of 72.6% (95% CI [69.72, 75.34]) with 6 observations per participant.

Sample.

One hundred and seventy-eight University of Toronto students completed the study in exchange for Introduction to Psychology course credit. We removed nine participants from our sample who, in generating six attitude objects, did not provide independent ratings for three disliked attitude objects and three hated attitude objects, resulting in a final sample size of 169 participants. We used repeated measures to increase within-subject power. The sample size of each laboratory study (Studies 1-3) was determined with the goal to optimize statistical power with the constraint of a convenient stopping point (e.g., running as many participants as possible until the subject pool closed at the end of the semester). We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study.

Design and Procedure.

Participants were asked to generate three disliked attitude objects and three hated attitude objects, in counterbalanced order. They rated these attitude objects on several dimensions and described the difference between these two sets of attitude objects.

Dependent Measures. Participants provided both quantitative and qualitative data for their six attitude objects. First, participants rated each of their attitude objects on 13 dimensions by conveying their level agreement with different statements on 7-point Likert scales (ranging from 1 = *strongly disagree* to 7 = *strongly agree*). These dimensions included valence (i.e. positivity: “Ignoring all my negative feelings, I feel very positive about this attitude object”; negativity: “Ignoring all my positive feelings, I feel very negative about this attitude object”), emotions (i.e. contempt: “I feel contempt toward this attitude object”; anger: “I feel anger toward this attitude object”; disgust: “I feel disgust toward this attitude object”; fear: “I feel fear toward this attitude object”, and annoyance: “I find the attitude object annoying/irritating”), attitude strength (i.e. certainty: “I feel very certain about my feelings toward this attitude object”; extremity: “I feel strongly about this attitude object”, and importance: “This attitude object is personally important to me”), motivation (i.e. approach: “I am motivated to confront this attitude object”, and avoidance: “I am motivated to avoid things related to this attitude object”), and morality (i.e. centrality to core moral beliefs).[1] Specifically, the morality item was adopted from Skitka and colleagues’ (Skitka et al., 2005) questionnaire: “My feelings about this attitude object are connected to my core moral beliefs or convictions.”. Finally, participants completed a free response item asking them to describe the similarities and differences between the three hated and three disliked attitude objects.

[1] The current paper focuses on the variables directly relevant to the morality and negativity hypotheses. Although other variables change from study-to-study, these variables remained consistent.

Analysis strategy. Analyses of repeated measures often focus on mean-level differences in ratings. However, this approach reduces multiple measures to a single score for each participant, diminishing power and meaningful variance. To analyze differences more accurately between hate and dislike, we used multilevel modeling (Hox, 1998). Multilevel modeling allows for the direct analysis of individual ratings and helps overcome violations of independence that occur due to correlated ratings within participants. When an assumption of independence is not satisfied, ignoring dependency among trials can lead to invalid statistical conclusions, namely, the underestimation of standard errors and the overestimation of the significance of predictors (Cohen et al., 2002). We therefore created multilevel models with repeated ratings nested within participants to provide more appropriate estimates of regression parameters. Multilevel models were implemented in R. We conducted a 2 (*Type*: Hate vs. Dislike) × 2 (*Order*: Hate first vs. Dislike First) mixed model design, between-subjects on the last factor. For completeness, we also report simple bivariate correlations of negativity, morality, and moral emotions for all three studies in Table 1.

Results

We began by examining the attitude objects participants generated. To visually illustrate these differences in an intuitive fashion, we created “Wordles” or world clouds, to visually represent the frequency with which each attitude object was listed. The wordles (see Figure 1) give greater prominence to words that are listed with the greatest frequency. The most frequently listed disliked attitude objects included “groups” and “rude.” The most frequently listed hated objects included “racism” and “tests.”

Quantitative Responses.

Intensity Hypothesis. To examine whether responses to hated attitude objects differ from disliked attitude objects in terms of degree of negativity, we first tested the intensity hypothesis, which predicts that hated objects are associated with greater negativity than dislike. Consistent with the intensity hypothesis, hated objects (*M* = 6.07, *SE* = .08) were rated more negatively than disliked objects (*M* = 5.74, *SE* = .08), M_{diff} = .33, *95% CI* [.20, .47], *t(825)* = 4.92, *p* < .001, *Cohen's d* = 0.23. We found no effect of order (i.e., hate first versus dislike first), *B* = -0.15, *95% CI* [-.47, .18], *t* (234) = -0.87, *p* = .39, nor did we find an interaction between attitude object type and the order in which participants generated their ratings, *B* = 0.03, 95% CI [-.24, .29], *t* (823) = 0.19, *p* = .84. This suggests that hated attitude objects are reported as more negative than disliked objects, consistent with the intensity hypothesis.

Effects of Intensity, adjusting for Morality. For a more stringent test of the intensity hypothesis, we also re-analyzed the effects of intensity adjusting for differences in morality. Given our mixed-model design, we adjusted for differences in morality in two ways. To begin, we computed a mean morality score for each participant by averaging each individual's six morality ratings for all of their attitude objects. Then, we created a between-person centered morality score to reflect how each participant's mean morality score differed from the average of all the participants' mean morality scores (i.e., the grand mean across all subjects). Second, we computed a within-person centered morality score to reflect how each trial differed from each individual's mean morality score across all six trials. Then, we re-analyzed the association between attitude object type (Hated vs. Disliked) and negativity, adjusting for both between-person and within-person centered morality scores. As a result, hated attitude objects were rated as more intensely negative than disliked attitude objects, M_{diff} = .28, *95% CI* [.15, .41], *t(822)* = 4.17, *p* < .001, even after adjusting for the effects of morality. Negativity ratings remained unaffected by the order in which participants generated them (*p* = .50). The effects of between-person morality, *B* = .16, *95% CI* [.01, .30], *t(163)* = 2.15, *p* = .033, and within-person morality, *B* = .13, *95% CI* [.09, .18], *t(821)* = 6.07, *p* < .001, were statistically significant in the model.

Morality Hypothesis. To examine whether hated objects differ from disliked objects in terms of morality, we next tested the morality hypothesis. Consistent with the morality hypothesis, hated objects (*M* = 5.33, *SE* = .10) were rated as more connected to morality than disliked objects (*M* = 4.91, *SE* = .10), M_{diff} = .42, *95% CI* [0.22, 0.63], *t(837)* = -3.99, *p* < .001, *Cohen's d* = 0.24. This effect was qualified by a significant interaction of attitude object type and order, *B* = .44, *95% CI* [.35, .92], *t(836)* = 2.10, *p* = .036, such that differences in morality between hated and disliked objects were significant for participants who were asked about the hated object first, M_{diff} = .63, *95% CI* [.35, .92], *t(835)* = 4.34, *p* < .001, *Cohen's d* = 0.34, but non-significant for participants who were asked about the disliked object first, M_{diff} = .19, *95% CI* [-.11, .49], *t(836)* = 1.26, *p* = .21, *Cohen's d* = 0.14, see Figure 2a. This suggests that hated attitude objects were perceived as more connected to morality than disliked objects, consistent with the morality hypothesis. However, rating hated attitude objects second resulted in attenuated differences in morality for hated versus disliked attitude objects.

We then examined whether hated attitude objects were associated with moral emotions—contempt, anger, and disgust—to a greater degree than disliked attitude objects. To do so, we created a moral emotions scale by averaging ratings of contempt, anger, and disgust for each participant (α = .68, see descriptive statistics of each emotion in Supplementary Table 1). Consistent with the morality hypothesis, hated objects (*M* = 4.95, *SE* = .08) were more strongly associated with moral emotions than disliked objects (*M* = 4.47, *SE* = .08), M_{diff} = .48, *95% CI* [0.35, .61], *t(838)* = 7.05, *p* < .001, *Cohen's d* = 0.37, see Figure 2b. We also found an effect of order, such that participants who generated hated attitude objects first reported experiencing less moral emotions (*M* = 4.52, *SE* = .10) than participants who generated disliked attitude objects first (*M* = 4.89, *SE* = .10), M_{diff} = .36, *95% CI* [0.35, .61], *t(167)* = 2.55, *p* = .012, *Cohen's d* = 0.28. The interaction of attitude object type and order of generation was not statistically significant, *B* = 0.06, *95% CI* [-0.21, 0.32], *t(838)* = 0.42, *p* = .67. This suggests that hated attitude objects were experienced as more closely connected to moral emotions than disliked objects, providing additional support for the morality hypothesis.

Effects of Morality, adjusting for Intensity. Given initial evidence for both the intensity and morality hypotheses, we examined the relationship between our main dependent variables. We found that morality ratings and the composite measure of moral emotions were highly correlated (*r* = .42, *p* < .001). This finding demonstrates construct validity by establishing the relation of moral judgments to affect that is central to morality (see Haidt, 2001). However, we also found that morality ratings and negativity were moderately correlated (r = .20, *p* < .001). This presents a potential confound for the intensity hypothesis and morality hypothesis (see Table 1). To help address this issue, we performed a more stringent test of the morality hypothesis by re-analyzing the data, adjusting for differences in negativity. Following the same strategy as presented above, we computed a mean negativity score for each participant by averaging each individual's negativity ratings for all of their attitude objects and created a between-person centered negativity score and a within-person centered negativity score. We then re-analyzed the degree to which attitude object type was associated with core moral beliefs, adjusting for both between-person and within-person centered negativity scores.

Providing further evidence for the morality hypothesis, hated attitude objects (*M* = 5.27, *SE* = .13) were rated as more connected to moral beliefs than disliked attitude objects (*M* = 4.74, *SE* = .13), M_{diff} = .54, *95% CI* [0.25, .82], *t(822)* = 3.72, *p* < .001, for participants who rated hated attitudinal objects first, even after statistically adjusting for the effects of intensity. However, differences in morality ratings were attenuated when hated attitude objects were rated second (*M* =4.73, *SE* = .13), M_{diff} = .08, *95% CI* [-0.22, .38], *t(823)* = 0.55, *p* = .58. The effect of between-person negativity, *B* = .20, *95% CI* [.04, .35], *t(169)* = 2.43, *p* = .016, and within-person negativity, *B* = .32, *95% CI* [.21, .42], *t(820)* = 5.96, *p* < .001, remained significant in the model. This pattern of findings provides evidence consistent with both the intensity hypothesis (that the difference between hate and dislike is the degree of negativity), and the morality hypothesis (such that above and beyond a difference in negativity, morality differentiates hate from dislike).

Similarly, we re-analyzed the data on our composite measure of moral emotions, adjusting for differences in negativity (i.e., both between- and within-person negativity). We again found an effect of hate versus dislike on the moral emotions, with hated attitude objects evoking more moral emotions (*M* = 4.90, *SE* = .08) than disliked attitude objects (*M* = 4.53, *SE* = .08), M_{diff} = .37, *95% CI* [0.24, .50], *t(824)* = 5.60, *p* < .001, even after adjusting for the effect of negativity. In this case, the effect of order became non-significant, *B* = -0.28, *95% CI* [-0.58, 0.01], *t(244)* = -1.87, *p* =.062. The effect of between-person negativity, *B* = .24, *95% CI* [.11, .38], *t(170)* = 3.47, *p* < .001, and within-person negativity, *B* = .32, *95% CI* [.25, .39], *t(823)* = 9.74, *p* < .001, remained significant in the model. Again, this pattern of findings supports both the intensity hypothesis, as well as the morality hypothesis.

Qualitative Responses

After reporting three disliked and three hated objects, participants were explicitly asked to compare their hated and disliked attitude objects. Two independent raters content coded participants' descriptions of the similarities and differences between their hated and disliked attitude objects and a third rater resolved any disagreements (Cohen's κ = .88). Only 103 participants answered this question; of that number, the raters agreed that 17 responses linked hated objects to morality, while only two responses linked disliked objects to morality. Compared to chance, where morality should be equally likely to be linked to either hate or dislike, we found that participants were more likely to report that hated objects were more closely linked to morality than disliked objects, $\chi2$ (1) = 11.84, $p < .001$. For instance, one participant wrote, "The hated objects were more moral issues in that they represent who I am and what my beliefs are. With those issues, I am prepared to take a stance and argue my point of view with the hope that others will agree. The disliked objects were less moral issues…These represent what appeals to my taste, but I'm not going to go out of my way to argue out loud or with uptight people". Another participant wrote, "The hated objects are more or less related to my inner beliefs and values of life whereas the other dislike objects are more…shallow aspects of my life". These results provide convergent evidence that morality plays a differentiating role between hate and dislike, lending further, qualitative support to the morality hypothesis.

Discussion

In our first study, we examined whether the conceptual differences between hate and dislike are simply a matter of degree of negativity or also a matter of morality. We found support for the *intensity* hypothesis—hated objects were viewed as more negative than disliked objects—suggesting that the difference between hate and dislike is indeed a matter of intensity. However, we also found support for the *morality* hypothesis—hated attitude objects were rated as more connected to participants' core moral beliefs and were associated with higher levels of moral emotions (contempt, anger, and disgust) than disliked attitude objects—suggesting that the difference between hate and dislike may also be a matter of morality. We found convergent evidence for this latter hypothesis across quantitative and qualitative analyses, with self-reports, expressions of moral emotions, and spontaneous descriptions.

We note that differences in morality were attenuated when participants were asked about disliked attitudinal objects first. We discuss possible explanations of this order effect below. Importantly, the results supporting the morality hypothesis remained significant even when adjusting for negativity. Above and beyond the effect of negativity, both moral concerns and moral emotions explained the variance in ratings of hated versus disliked attitude objects. Likewise, participants spontaneously reported that hated objects were more closely tied to morality than disliked objects in their qualitative responses. These findings provide preliminary evidence that the conceptualization of hate may differ from dislike, and that morality may play a key role in explaining this difference.

We also found unexpected effects due to the order in which participants generated hated versus disliked attitude objects. The main effect of hate versus dislike on morality ratings was qualified by an interaction in which the order of generating the attitude objects moderated morality judgments. In particular, morality ratings were lower for disliked attitude objects when they were generated after hated ones. One potential explanation is that when participants thought about hate first (then dislike), they understood the nature of the conceptual contrast between hate versus dislike and attenuated their responses for their disliked attitude objects. However, when participants were asked to generate disliked attitude objects first, they may have reported things they hate, and only later understood the contrast once they were invited to list hated objects. Indeed, previous research finds that nearly a third of people consider hate to be synonymous with extreme dislike (Aumer-Ryan & Hatfield, 2007).

In reviewing participants' responses in the free response section, we found several comments that support this interpretation of our data. For instance, one participant remarked: "I was given the 'dislike' paper first and then the 'hate' paper. Hence, I did not understand that I had to put emphasis on the word 'dislike.' I made a little mess in that," and another wrote "If I had known I was going to be writing about things I 'hated' after the things I 'disliked' I would have had some of the 'disliked' responses be hated responses." Given that some participants generated hated attitude objects in the dislike condition, our results likely represent a conservative test of the differences between hate and dislike. Although this speaks to the semantic challenges of studying this issue, it is important to note that participants reported more moral emotions for hated attitude objects—regardless of order. Thus, any effects of order cannot account for the overall pattern of results.

Study 2: Hate versus Extreme Dislike

The findings in Study 1 provided the first empirical evidence that morality differentiates between hated and disliked attitude objects, even after adjusting for differences in the degree of negativity. However, given that participants attenuated their morality ratings for disliked attitude objects when they rated them second (i.e., after seeing an explicit contrast between hate and dislike), we sought to address the observed but unpredicted order effects in Study 2. First, participants were made aware in the instructions that they would be generating both hated and disliked attitude objects. Second, all participants generated their attitude objects in the same order. Third, to help disguise our main research question and minimize demand effect, we also had participants rate liked and loved objects and complete a number of irrelevant emotion ratings. Fourth, we also included a condition in which participants compared hate to extreme dislike. This new condition offered a more stringent test of the hypotheses that hated and disliked objects differentiate in morality by comparing hated objects directly with extremely disliked objects. If morality does differentiate between hate and dislike, hated attitude objects should be rated as more connected to moral concerns and as evoking moral emotions to a greater degree than both disliked and extremely disliked attitude objects. We also added manipulation checks for both hate and dislike. Finally, we collected data for Studies 2 and 3 in another country (USA) to increase the diversity of our samples and see if the results would generalize outside of Canada.

Method

For study 2, the power analysis based on 1000 simulations and an alpha = 0.05 showed that recruiting 176 participants would enable the detection of small effects (e.g., 0.3) for the interaction between within-subjects attitude object type (hated vs. disliked object) and between-subjects condition (extreme vs. regular) with statistical power of 81.3% (95% CI [78.74, 83.67]) with 2 observations per participant.

Sample.

One hundred and eighty-two The Ohio State University students completed the study in exchange for Introduction to Psychology course credit. We removed five participants from our sample who entered only one or two scale rating numbers (of seven possible ratings) for all of their responses to the dependent measures in the study (suggesting that they were not paying any attention or engaged in the study) resulting in a final sample of 177 participants. Again, we used a within-subjects contrast to increase power.

Design and Procedure

Participants were asked to generate one hated, one disliked, one liked, and one loved attitude object in that order. Participants were randomly assigned to either a regular attitude objects condition (similar to Study 1), or to an extreme attitude objects condition in which they generated attitude objects for hate, extreme dislike, extreme like, and love.

Dependent Measures. Participants provided quantitative data for their four attitude objects. As in Study, 1, participants rated each of their attitude objects on the same 13 dimensions (i.e., valence, emotions, attitude strength, motivation, and morality). To help disguise our intent, we included several additional attitudinal (e.g., feel strongly, important to me) and emotional (e.g., distressed, conflicted, excited) dimensions from the expanded Positive and Negative Affect Schedule (PANAS-X; Watson & Clark, 1994). Thus, participants completed a total of 40 ratings per attitude object. We also added two manipulation check items—how much participants felt hate and dislike toward their attitude objects—which were mixed in with the PANAS-X items to disguise the intent of our manipulation checks. Like in Study 1, all ratings were made on 7-point Likert scales (ranging from 1 = *strongly disagree* to 7 = *strongly agree*).

Analysis. We focus on responses for the hated and disliked (or extremely disliked) attitude objects, so no analyses for the liked and loved attitude objects are presented. As in Study 1, we created multilevel models with repeated ratings nested within participants to provide more appropriate estimates of regression parameters. We conducted a 2 (*Type*: Hate vs. Dislike) × 2 (*Extremity*: Regular dislike vs. Extreme Dislike) mixed model design, between-subjects on the last factor.

Results

Manipulation Checks

We included ratings of dislike and hate as manipulation check items, mixed in with PANAS-X items. We wanted to confirm that participants hated their hated attitude objects more than their disliked attitude objects. Regarding our dislike manipulation check, we were more agnostic in our predictions. Participants might dislike their hated attitude objects more than their disliked ones or that they could dislike both attitude objects equally.

Hated attitude objects were more disliked (*M* = 4.21, *SE* = .13) than disliked attitude objects (*M* = 3.67, *SE* = .13), M_{diff} = .54, 95% CI [0.24, 0.84], *t(171)* = 3,55, *p* < .001, *Cohen's d* = 0.42, but only in the regular condition. In the extreme condition, dislike ratings did not differ based on attitude object type, M_{diff} = -0.02, 95% CI [-0.33, 0.28], *t(171)* = -0.15, *p* = .88, *Cohen's d* = -0.01. Our dislike manipulation check showed that hated attitude objects were more disliked than disliked attitude objects, but that hated attitude objects were equally disliked as extremely disliked attitude objects.

As predicted, hated attitude objects were more hated (*M* = 4.05, *SE* = .10) than disliked attitude objects (*M* = 3.12, *SE* = .10), *B* = 0.92, 95% CI [0.67, 1.17], *t(171)* = 7.31, *p* < .001, *Cohen's d* = 0.64. This effect was qualified by a significant interaction of attitude object type and extremity condition, *B* = -0.63, 95% CI [-0.33, 0.28], *t(171)* = -2.52, *p* = .012. In the regular condition, hate ratings for hated attitude objects (*M* = 4.16, *SE* = .15) were greater than for disliked attitude objects (*M* = 2.92, *SE* = .15), M_{diff} = 1.24, 95% CI [0.89, 1.59], *t(171)* = 6.98, *p* < .001, *Cohen's d* = 0.94. In the extreme condition, hate ratings for hated attitude objects (*M* = 3.93, *SE* = .15) were still significantly greater, but relatively less so, than for disliked attitude objects (*M* = 3.33, *SE* = .15), M_{diff} = 0.60, 95% CI [0.25, 0.96], *t(171)* = 3.38, *p* < .001, *Cohen's d* = 0.37. Our hate manipulation check revealed that hated attitude objects were more hated than disliked attitude objects, but especially in the regular dislike condition. Thus, while hate and extreme dislike were equally negative, hate was still rated as conceptually distinct from extreme dislike.

Intensity Hypothesis.

We examined the intensity hypothesis, which predicts that the difference between hated and disliked objects is a matter of degree of negativity. We replicated our effect of hate versus dislike, such that hated attitude objects were rated more negatively (*M* = 5.76, *SE* = .13) than disliked ones (*M* = 5.44, *SE* = .13), M_{iffs} = 0.32, 95% CI [0.03, 0.61], *t(171)* = 2.17, *p* = .03, *Cohen's d* = 0.19. We found no effect of extremity condition and no interaction of attitude object type and extremity condition (*ps* > .41). This suggests that hated attitude objects are perceived to be more negative than disliked ones, replicating our findings from Study 1 supporting the intensity hypothesis.

Effects of Intensity, adjusting for Morality. We re-analyzed the effects of intensity adjusting for both between-person and within-person centered morality scores. Hated attitude objects were rated as more intensely negative than disliked attitude objects, M_{diff} = 0.34, 95% CI [0.04, 0.63], *t(170)* = 2.21, *p* = .03, even after adjusting for the effects of morality. Negativity ratings remained unaffected by the extremity condition in which participants generated them (*p* = .44). The effects of between-person morality and within-person morality were non-significant in the model (*ps* < .15).

Morality Hypothesis.

Replicating the results of Study 1, hated attitude objects were rated as more connected to core moral beliefs (*M* = 5.20, *SE* = .15) than disliked ones (*M* = 4.60, *SE* = .15), M_{diff} = 0.60, 95% CI [0.26, 0.94], *t(171)* = 3.45, *p* < .001, *Cohen's d* = 0.31, see Figure 3a. We found no effect of extremity condition and no interaction of attitude object type and extremity condition (*ps* > .19). Further analysis revealed that the simple effect of hate versus dislike in the extreme condition was not only statistically significant, but, if anything, actually greater than in the regular condition, M_{diff} = 0.83, 95% CI [0.34, 1.31], *t(171)* = 3.37, *p* < .001, *Cohen's d* = 0.43.

This suggests that hated attitude objects are perceived as more connected to morality than disliked attitude objects—especially when hated attitude objects are compared to extremely disliked attitude objects.

We further tested the morality hypothesis by investigating the relationship between hate versus dislike on a composite measure of moral emotions (i.e., contempt, anger, and disgust, α = .62). Replicating the results of Study 1, hated attitude objects evoked a greater degree of moral emotions (*M* = 3.28 *SE* = .08) than disliked attitude objects (*M* = 2.89, *SE* = .08), M_{diff} = 0.39, 95% CI [0.22, 0.56], *t(171)* = 4.52, *p* < .001, *Cohen's d* = 0.37, see Figure 3b. Again, we found no effect of extremity condition and no interaction of attitude object type and extremity condition (*ps* > .17). This finding provides additional support for the morality hypothesis by showing that hated attitude objects are associated with a greater experience of moral emotions than disliked attitude objects.

Effects of Morality, adjusting for Intensity

Given that we replicated our findings in support of the hypotheses that the difference between hate and dislike is both a matter of intensity and a matter of morality, we tested the inter-correlations of our main dependent variables. As in Study 1, we found that morality judgments and moral emotions were positively correlated (*r* = .29, *p* < .01) and that negativity and the moral emotions were moderately correlated (*r* = .23, *p* < .01), but negativity and morality were not significantly correlated (*r* = .075, *p* = .16) (see Table 1). To perform a more stringent test of the morality hypothesis, we re-analyzed the effects of attitude object type and extremity condition on moral judgments and the moral emotions, adjusting for both between-person and within-person negativity.

We replicated our Study 1 findings of the effect of hate (*M* = 5.20, *SE* = .15) versus dislike (*M* = 4.59, *SE* = .15) on morality, M_{diff} = 0.61, 95% CI [0.26, 0.96], *t(170)* = 3.47, *p* < .001, even when adjusting for negativity in two ways. We found no effect of extremity condition, between-person negativity, or within-person negativity, and no interaction of attitude object type and extremity condition (*ps* > .15). In Study 1, both between and within-person negativity remained significant predictors of attitude object type (i.e., hate versus dislike), adjusting for morality. However, in Study 2, attitude object type (i.e., hate versus dislike) was the only significant predictor of morality ratings. This pattern of findings supports the morality hypothesis, such that above and beyond a difference in negativity, morality differentiates hate from dislike.

Similarly, even when adjusting for between-person and within-person negativity, we replicated the effect of hate (*M* = 3.26, *SE* = .08) versus dislike (*M* = 2.91, *SE* = .08) on the expression of moral emotions, M_{diff} = 0.35, 95% CI [0.18, 0.52], *t(170)* = 4.09, *p* < .001. We found no effect of extremity condition and no interaction of attitude object type and extremity condition (*ps* > .19). As in Study 1, the effect of between-person negativity, *B* = 0.13, 95% CI [0.04, 0.22], *t(170)* = 2.94, *p* = .003, and within-person negativity, *B* = 0.13, 95% CI [0.04, 0.21], *t(170)* = 2.85, *p* = .005, remained significant predictors of morality, adjusting for attitude object type. This pattern of findings similarly supports the intensity hypothesis as well as the morality hypothesis, such that above and beyond a difference in negativity, the experience of moral emotions differentiates hate from dislike.

Discussion

In Study 2, people were asked to generate hated, disliked, liked, and loved attitude objects. By examining the contrasts of hate versus dislike and hate versus extreme dislike, we were able to explore a more stringent test of our morality hypothesis. Replicating the results of Study 1, we found additional support for the *intensity hypothesis* suggesting that the difference between hate and dislike conceptualizations is indeed a matter of degree of negativity. We also found strong support for the *morality hypothesis*, finding that attitude objects were rated as more connected to peoples' core beliefs and were associated with moral emotions (i.e., contempt, anger, and disgust) to a greater degree than disliked attitude objects. Strikingly, when people contrasted between hate versus extreme dislike, they rated their hated attitude objects as even *more* moral than when hate was contrasted with regular dislike. In addition, the results supporting the morality hypothesis again remained significant even when adjusting for between-person and within-person negativity. This means that above and beyond the effect of negativity, morality (and moral emotions) differentiated between hated and disliked attitude objects.

Study 3: Hate, Moral Conviction, & Moral Concern

Thus far, we have investigated the relationship between hate versus dislike on ratings of negativity, morality, and the moral emotions. However, research on moral convictions (or moral mandates) describes further ways in which morality may be defined and measured (see Skitka, 2002; Skitka et al., 2005; Skitka & Mullen, 2002). Specifically, moral convictions may be distinguished from otherwise strong but non-moral attitudes by the experience of a unique combination of universalism, factual belief, and justification for action (Skitka et al., 2005). According to this account, moral convictions are experienced as objective characteristics of the world (factual belief) that are perceived as absolutes (universalism), disregarding cultural differences or personal preferences (Haidt et al., 2003), and motivate action (justification for action) (Skitka et al., 2005). These aspects of morality have been shown to predict action potential to a greater extent than non-moral attitudes even after controlling for attitude strength (e.g., extremity, importance, and certainty). For this reason, in Study 3 we included these additional indices of moral conviction (universalism, factual belief, and justification for action) to investigate differences in morality between hate and dislike and controlled for differences in attitude strength.

If differentiating hated from disliked objects is a matter of intensity, hated attitude objects should be rated as more negative than both disliked and extremely disliked attitude objects. Further, if morality does play a differentiating role between hated and disliked objects, hated attitude objects should be rated as more connected to morality (including universalism, factual belief, and justification for action) than both disliked and extremely disliked attitude objects, even after controlling for attitude strength.

In addition, there remains a potential alternative explanation for our pattern of results in Studies 1 and 2: perhaps asking participants to produce examples for hate versus dislike (or extreme dislike) evokes responses that reflect different classes of attitude objects. To help ensure this was not the case, Study 3 sought to replicate and extend the differences in morality and intensity observed in Studies 1 and 2 when *assigning* participants to generate specific classes of attitudes objects: either people/groups or concepts/beliefs.

Method

For study 3, a power analysis based on 1000 simulations and an alpha = 0.05 revealed that recruiting 78 participants would enable the detection of small effects (e.g., 0.3) for the interaction between within-subjects attitude object type (hated vs. extremely disliked vs. disliked object) and between-subjects attitude object class (concepts/beliefs vs. people/groups) with a statistical power of 87.1% (95% CI [75.14, 78.88]) with 3 observations per participant.

Sample.

Eighty-two students at The Ohio State University completed the study in exchange for Introduction to Psychology course credit. We removed four participants from our sample that reported “nothing” or “no one” for their hated attitude object (i.e., failed to report something they hated), resulting in a total sample of 78 participants. Again, we used a within-subjects contrast to increase power.

Design and Procedure.

Participants were asked to generate one disliked, one extremely disliked, and one hated attitude object in that order. Participants were either assigned to a people/groups attitude objects condition or to concepts/beliefs attitude objects condition. In each condition, participants were instructed to list either a person/group or concept/belief they hated (e.g., “For this phase of the study we want you to list a PERSON or GROUP you HATE.”).

Dependent Measures. Participants provided quantitative data for their three attitude objects. As in Studies 1 and 2, participants rated each of their attitude objects on the same 13 dimensions (i.e., valence, emotions, attitude strength, motivation, and morality), as well as rated the other attitudinal (e.g., feel strongly, important to me) and emotional (e.g., distressed, conflicted, excited) dimensions from Study 2. In addition, participants provided ratings on items related to *morality*, including the same item employed in Study 1 and 2 (“My feelings about this attitude object are connected to my core moral beliefs or convictions”), and eight additional morality items (i.e. “____violates my core moral beliefs”, “____intentionally violates my core moral beliefs”, “____is saintly” (reverse-coded), “____is evil”, “____has the right core moral beliefs” (reverse-coded), “____has the wrong core moral beliefs”, “____has no core moral beliefs”, and “every time I think of ____ my core moral beliefs spring to mind”), as well as items related to *universality* (“Any reasonable person would share my feelings about ____”, and

"Feelings about ____ are a matter of personal taste" (reverse-coded)), *factual belief* ("it's a fact that ____ is wrong" and "it's a fact that ____ is right" (reverse coded)), and *justification for action* ("I feel morally obligated to do something about____" and "If I took action against ____, it would validate my moral beliefs") for a total of 51 ratings per attitude object. Like in the previous studies, all ratings were made on 7-point Likert scales ranging from "1 = *strongly disagree*" to "7 = s*trongly agree*."

To test the relationship between hate and morality, we created a morality scale by analyzing all morality items (the items employed in Studies 1 and 2, and the eight additional items described above), as well as the items on universalism, factual belief, and justification for action (following Skitka et al., 2005). We completed factor analyses on these fifteen items and eliminated the items with loadings below 0.299 ("saintly" (reverse-coded) and one of the *factual belief* items, "feelings about ____ are a matter of personal taste" (reverse-coded)), resulting in a 13-item moral conviction scale with very good reliability α = .87). Thus, the final moral conviction scale included most of the items related to *universality, factual belief,* and *justification for action*.

Analysis. As in Studies 1 and 2, we created multilevel models with repeated ratings nested within participants to provide more appropriate estimates of regression parameters. Multilevel models were implemented in R. We conducted a 3 (*Type*: Hate vs. Extreme Dislike vs. Dislike) × 2 (*Attitude Object Class*: People/Groups vs. Concepts/Beliefs) mixed model design, between-subjects on the last factor.

Results

Intensity Hypothesis.

Replicating the results of Studies 1 and 2, attitude object type had a significant effect on negativity, but only for some of the contrasts (Hate vs. Dislike: *B* = 0.77, *95% CI* [-0.39, 1.14], *t(154)* = 4.02, *p* < .001, but not Hate vs. Extreme Dislike: *B* = 0.23, *95% CI* [-0.14, 0.61], *t(154)* = 1.21, *p* = .23). In particular, hated attitude objects (*M* = 5.99, *SE* = .15*)* were rated more negatively than disliked attitude objects (*M* = 5.21, *SE* = .15*)*, M_{diff} = 0.78, *95% CI* [0.33, 1.24], *t(152)* = 4.10, *p* < .001, *Cohen's d* = 0.51, see Figure 4a. However, hated attitude objects were not significantly different from extremely disliked attitude objects (*M* = 5.76, *SE* = .15), M_{diff} = 0.24, *95% CI* [-0.22, 0.69], *t(152)* = 1.23, *p* = .44, *Cohen's d* = 0.09. Extremely disliked attitude objects were rated more negatively than disliked attitude objects, M_{diff} = 0.55, *95% CI* [0.10, 1.00], *t(152)* = 2.87, *p* = .013, *Cohen's d* = 0.42.

We also found a main effect of attitude object class on ratings of negativity, such that people/groups (*M* = 5.86, *SE* = .14) were rated more negatively than concepts/beliefs (*M* = 5.45, *SE* = .13), M_{diff} = 0.41, *95% CI* [0.03, 0.79], *t(76)* = 2.14, *p* = .04, *Cohen's d* = 0.32. We found no interaction of hate versus dislike and attitude object class (*ps* > .12). This pattern of results suggests that hated attitude objects are perceived to be more negative than disliked ones, replicating our findings from Studies 1 and 2 in support of the intensity hypothesis. Importantly, we did not find that hated attitude objects are perceived to be more negative than extremely disliked attitude objects.

Effects of Intensity, adjusting for Morality.

We re-analyzed the effects of intensity adjusting for both between-person and within-person centered morality scores. After adjusting for the effects of morality, differences in negativity became non-significant across all contrasts: hated (*M* = 5.65, *SE* = 0.13) versus disliked (*M* = 5.57, *SE* = 0.13) attitude objects, M_{diff} = 0.08, *95% CI* [-0.35, 0.52], *t(151)* = 0.44, *p* = .90, hated versus extremely disliked (*M* = 5.72, *SE* = 0.12) attitude objects, M_{diff} = -0.06, *95% CI* [-0.25, 0.55], *t(151)* = -0.41, *p* = .91, and disliked versus extremely disliked attitude objects, M_{diff} = 0.15, *95% CI* [0.04, 0.63], *t(151)* = 0.88, *p* = .66. The effects of between-person morality, *B* = 0.72, *95% CI* [0.47, 0.96], *t(151)* = 5.70, *p* < .001, and within-person morality, *B* = 0.72, *95% CI* [0.55, 0.90], *t(75)* = 7.99, *p* < .001, were statistically significant in the model, while object class (concept/belief versus person/group) and its interaction with attitude object type were non-significant (*ps* > 12). Thus, the intensity hypothesis did not survive the inclusion of morality.

Morality Hypothesis

Replicating the results from Study 1 and 2 with regard to the relation between hate and morality, attitude object type significantly predicted ratings of morality (Hate vs. Extreme Dislike: *B* = 0.41, *95% CI* [0.13, 0.70], *t(154)* = 2.86, *p* = .005, and Hate vs. Dislike: *B* = 0.97, *95% CI* [0.68, 1.25], *t(154)* = 6.69, *p* < .001). Specifically, hated attitude objects (*M* = 5.25, *SE* = .11) were more related to morality than disliked attitude objects (*M* = 4.27, *SE* = .11), M_{diff} = 0.98, *95% CI* [0.04, 0.63], *t(152)* = 6.75, *p* < .001, *Cohen's d* = 0.90, or extremely disliked attitude objects (*M* = 4.82, *SE* = .11), M_{diff} = 0.42, *95% CI* [0.04, 0.63], *t(152)* = 2.90, *p* = .012, *Cohen's d* = 0.37, see Figure 4b. Extremely disliked attitude objects were also more related to morality than disliked ones, M_{diff} = 0.56, *95% CI* [0.04, 0.63], *t(152)* = 3.85, *p* < .001, *Cohen's d* = 0.49. We found no effect of attitude object class and no interaction of hate versus dislike and attitude object class (*ps* > .08). These findings further support a distinction between hated and disliked attitudes objects rooted in morality[2].

Effects of Morality, adjusting for Intensity and Attitude Strength.

[2] In this study, a computer error made it impossible to obtain moral emotion items (i.e. contempt, anger, and disgust) as well as several of the PANAS-X items failed for the majority of participants. For this reason, we report analyses for moral concern and moral conviction as opposed to the moral emotion analyses presented in Studies 1 and 2.

Consistent with hypotheses that the difference between hated and disliked attitude objects is both a matter of intensity and a matter of morality, we tested the inter-correlations of our main dependent variables. Morality and negativity were highly significantly correlated (*r* = .59, *p* < .001) (see Table 1). To perform a more stringent test of the morality hypothesis, we re-analyzed the effects of attitude object type and attitude object class on morality, adjusting for both between-person and within-person negativity. Even after controlling for between and within-person negativity, hated attitude objects (*M* = 5.10, *SE* = .10) were more related to morality than disliked attitude objects (*M* = 4.45, *SE* = .10), M_{diff} = 0.65, *95% CI* [0.35, 0.96], *t(151)* = 5.09, *p* < .001, or extremely disliked attitude objects (*M* = 4.78, *SE* = .09), M_{diff} = 0.32, *95% CI* [0.03, 0.61], *t(151)* = 2.64, *p* = .025. Extremely disliked attitude objects were more related to morality than disliked ones, M_{diff} = 0.33, *95% CI* [0.03, 0.63], *t(151)* = 2.64, *p* = .025. We found no effect of attitude object class and no interaction of hate versus dislike and attitude object class (*ps* > .39). The effect of between-person negativity, *B* = 0.42, *95% CI* [0.29, 0.57], *t(75)* = 5.70, *p* < .001, and within-person negativity, *B* = 0.41, *95% CI* [0.31, 0.51], *t(151)* = 7.99, *p* < .001, also remained significant in the model.

Morality and attitude strength were significantly correlated (r = 0.57, p < .001). Thus, we repeated the same procedure to evaluate the effects of morality after controlling for attitude strength. Hated attitude objects were still more associated to morality (*M* = 5.10, *SE* = .10) than disliked attitude objects (*M* = 4.42, *SE* = .10) after controlling for attitude strength, M_{diff} = 0.68, *95% CI* [0.36, 1.00], *t(151)* = 4.97, *p* < .001. However, the contrast between hated and extremely disliked attitude objects (*M* = 4.82, *SE* = .10) became only marginally significant, M_{diff} = 0.28, *95% CI* [-0.03, 0.59], *t(151)* = 2.14, *p* = .085. The effects of between-person attitude strength, *B*

= .68, *95% CI* [.50, .87], *t (75)* = 7.22, *p* < .001, and within-person attitude strength, *B* = .47, *95% CI* [.32, .60], *t(151)* = 6.40, *p* < .001, were significant.

When evaluating differences in attitude strength, hated attitude objects were associated with increased attitude strength (*M* = 4.96, *SE* = 0.11) compared to disliked objects (*M* = 4.32, *SE* = 0.11), M_{diff} = 0.64, *95% CI* [0.30, 0.98], *t(152)* = 4.44, *p* < .001, *Cohen's d* = 0.56, though no differences were found between hated and extremely disliked objects in terms of attitude strength (*M* = 4.66, *SE* = 0.11), M_{diff} = 0.30, *95% CI* [-0.04, 0.64], *t(152)* = 2.09, *p* = .095, *Cohen's d* = 0.28, see Figure 4c. Moreover, after adjusting for between and within-participants' morality, even differences in attitude strength between hated compared to disliked attitude objects became non-significant (*p* = .39).

Because differences in negativity did not survive adjusting for morality, the findings from Study 3 support the morality hypothesis but not the intensity hypothesis. Moreover, differences in morality between hated and disliked attitude objects remained significant even when adjusting for between-person and within-person negativity and attitude strength. Of note, attitude strength was similar between hated and extremely disliked attitude objects and morality differences between hated and extremely disliked attitude objects became non-significant after statistically adjusting for attitude strength.

Discussion

This study further tested whether the difference between hated and disliked attitude objects is one of intensity or morality. As in the previous studies, we found that hated objects were rated more negatively than disliked attitude objects, supporting the *intensity hypothesis.* However, hated objects did not differ from extremely disliked objects in terms of negativity and differences in negativity did not survive controlling for morality, contradicting the idea that hate simply falls at the extreme end of a continuum of negativity. Rather, we found evidence that hate is more connected to moral convictions when compared to disliked attitude objects—even when adjusting for negativity and attitude strength. Moreover, hated objects were also more related to morality than extremely disliked objects after controlling for negativity, though these differences were only marginally significant after adjusting for attitude strength. Because attitude strength ratings were similar for hated and extremely disliked attitude objects, attitude strength cannot account for differences between hate and extremely disliked attitude objects. Thus, hate appears to be different from dislike in terms of morality rather than negativity or attitude strength.

Study 3 also examined the types of attitude objects participants evaluated: either people and groups or concepts and beliefs. Whereas people and groups were rated more negatively than concepts and beliefs, they did not differ along the moral dimension, suggesting that all hated attitude objects share moral relevance. This suggests that our conclusions about differences in morality generalize to multiple targets of hate. Although we selected this distinction based on the types of attitude objects generated in the first two studies, it is possible that such instructions led to the selection of less central, more socially appropriate attitude objects. To test this concern and see if our findings would generalize to real world hate groups, we examined differences in the online expression of hate versus dislike.

Study 4: Hate Online

In the fourth study, we investigated real world-instantiations of hate versus dislike expressions on the Internet. Online media have been a prominent platform for transmitting messages of hate—even in 1995, in the first days of the commercialized Internet, hate groups were some of the earliest to leverage this technology to recruit members, organize, and transmit their beliefs. For instance, Stormfront.org is one of the world's oldest hate groups—providing an

Internet forum for over 300,000 registered users who are affiliated with neo-Nazi and White supremacist groups. This website is run by a former Ku Klux Klan leader and has been accused of promoting deadly violence, with connections to nearly 100 killings (Holpuch, 2014). To date, hate groups flourish online and use the Internet and social media to mobilize collective action (Hoover et al., 2021).

We analyzed the linguistic content of websites belonging to real, American hate groups as well as complaint forums (both consumer complaints and forums expressing employee complaints against corporations), in order to determine if the language used on these websites reveals whether they differ in their level of negativity or in their structure as characterized by morality. Because hate groups often do not identify with the label “hate group” themselves, we use a conceptualization of hate defined by how others see these groups rather than how they see themselves, unlike in Studies 1 to 3. Including an online sample also allowed us to capture a diversity of perspectives not reflected among our previous undergraduate samples.

Method

Sample.

We sampled websites, including 46 sites of known hate groups (expressions of hate), 47 threads from different subcategories of an online consumer complaint forum (expressions of dislike towards products/objects), and 51 threads of employee complaints about different corporations (expressions of dislike towards people/corporations). Hate groups were selected from the Southern Poverty Law Center’s Hate Map and chosen because they were among the most commonly noted groups among America’s 50 states that had publicly accessible websites (and were not specifically noted as “separatist groups” or religious fundamentalist groups[3]).

[3] See osf.io/5u6my for a full list of sites and text files.

To serve as one natural comparison condition, consumer complaint forums were selected from subcategories of the international complaint site, complaintsboard.com.[4] This maintained consistency in the format of the different complaint threads. Employee complaints about corporations were selected from Glassdoor.com, one of the most widely used corporate review sites, which verifies posters and screens content for accuracy, assuring as much consistency as possible across types of posts (Associated Press, 2013). We focused on the 51 companies with the greatest number of reviews that also had enough 2-star or lower reviews for us to cull to match the word count from the hate group and complaint forum texts.[5] Taken together, we could compare sites in which hate is expressed against sites in which dislike is expressed (against objects or corporate groups).

Design & Procedure

As all hate websites are different, in selecting the text for our sample we tried to maintain as much consistency as possible within and between websites. From the hate websites we analyzed the text from their “About” and “Mission Statement” sections. We then analyzed a sufficient number of complaint threads (both from Complaintsboard and Glassdoor) per subcategory or company to match the word length of hate websites. To compare the use of moral words, we manually counted usages of the word “moral” and its synonyms.[6] For example, one such usage by the hate group, Gallows Tree Wotansvolk declared:

[4] At the time of data collection, August 17, 2017, www.complaintsboard.com was the #1 hit upon a Google search of “complaint forums.”

[5] Data was collected from Glassdoor in September 2019.

[6] We searched for the synonyms given by the Merriam Webster Dictionary—conscionable, ethical, honest, honorable, just, principled, scrupulous—in their noun, adjectival, and adverbial forms. Antonyms (e.g. unethical, injustice) were included in this search; however, alternate meanings were not included (e.g. when “just” is taken to mean exactly or precisely). Additionally, spelling errors were considered—when “moral” was meant to represent “morale” it was not included, but when “morale” was, by context, intended to mean “moral” it was counted.

"Votanism teaches lessons of ***morality*** *and nobility, to walk as a proud White individual in a world where being White is now considered wrong…We understand honor to be one of the foundation blocks which will support our healthy growth and advancement. Thus, we wish to fill our ranks with men and women of* ***honor****..."*

One example of moral word usage by a complaint site was featured on the Consumer Electronics complaint forum on Complaintsboard.com:

"First they cheat customers by selling refurbished and used products in the original packaging. Second, they have made a joke out of customers by being ***unethical*** *and unresponsive."*

Finally, to search for differences in word length as well as the emotional content of words, we employed the Linguistic Inquiry and Word Count program (Pennebaker, Francis, & Booth, 2003). We compared word frequencies, expressed as a percentage of the total word count, between hate sites and complaint forums.

Results

We first analyzed our data to determine that our hate sites and two forms of complaint sites did not fundamentally differ in number of words. A one-way ANOVA revealed that the text taken from the hate websites (*M* = 1279.65 *SD* = 1731.59) did not significantly differ from the consumer complaint forums (*M* = 1434.89, *SD* = 125.65) nor the employee complaints (*M* = 1273.51, *SD* = 352.96) in terms of the average overall word count, *F*(2,141) = .40, *p* = .67, η^2_p =.006. As such, any linguistic differences noted between the sites were not due to differences in composition length.

Intensity Hypothesis. We explored whether the different types of websites used different frequencies of negative words, as a test of the intensity hypothesis. The LIWC dictionary examines negative language by comparing the usage of approximately 495 different words related to negative emotion, including “annoyed,” “grief,” and “hurt,” and then provides the percentage of negative words per hate website (or complaint threads with matched word length). We specially edited the LIWC dictionary to remove synonyms for the moral emotions (e.g., anger, contempt, disgust) as well as words related to morality (e.g., steal, guilt, punish), to ensure that we were examining differences in negativity separate from moral concerns. More frequent use of negative emotion words on the part of one type of site would lend support for the intensity hypothesis. We found that the amount of negative words used differed across sites, $F(2,141) = 23.72$, $p <.001$, $\eta^2_p =.25$. Specifically, hate websites ($M = 1.31$, $SD = .85$) did not use more negative words than consumer complaint threads ($M = 1.40$, $SD = .50$), $t(141) = .70$, $p = .49$. However, employee complaints featured more negative words ($M = 2.12$, $SD = .53$) than both hate sites, $t(141) = 6.24$, $p = <.001$, and consumer complaints, $t(141) = 5.56$, $p = <.001$. This suggests that hate sites do not evince more negative emotion words than complaint sites and may even use less negative emotion words than employee complaints. Thus, the intensity hypothesis was not supported in this context of real online expressions of hate and dislike.

Morality Hypothesis. We next explored the morality hypothesis by comparing the usage of the word “moral” and its synonyms across sites. There was a significant effect of site type predicting moral word usage, $F(2,141) = 11.35$, $p < .001$, $\eta^2_p =.14$. Replicating our three lab studies, the rhetoric on hate websites ($M = 0.32$, $SD = 0.50$) contained significantly more moral words than consumer complaint forums ($M = 0.04$, $SD = 0.07$), $t(141) = 4.57$, $p < .001$, and significantly more moral words than employee complaint forums ($M = 0.11$, $SD = .311$), $t(141) = 3.51$, $p < .001$. Further, this pattern of results persisted when accounting for the variability in negative emotion word usage, $F(2, 140) = 10.11$, $p < .001$, $\eta^2_p =.13$. Thus, our analysis of close to

150 online websites and forum threads suggests that hate groups use more moral language than complaint forums.

Discussion

Consistent with our three laboratory studies, this analysis of real-world hate groups lent further support to the morality hypothesis. The language of hate groups was different, in the moral domain, than that of complain forums. These results reflect real, uncensored language used by groups and individuals known to espouse hate or dislike, and often known to take significant actions in support of these attitudes. Importantly, the expressions of these groups and individuals are less likely to involve their own lay theories about hate or dislike—but rather their public positions on these issues. There are, admittedly, several differences between hate websites and complaint forums and these data should be treated as preliminary. But taken together with the carefully controlled lab experiments, this overall pattern of findings suggests that morality helps differentiate expressions of hate from expressions of dislike.

General Discussion

In a combination of laboratory studies and a content analysis of real online hate and complaint websites, we found initial evidence that differences in people's conceptualizations of hate and dislike are not only a matter of negativity but also morality. Morality—both via the expression of moral emotions and moral conviction—differentiates hated from disliked attitude objects. Individuals rated hated attitude objects in the lab as more closely connected to morality than disliked or even extremely disliked attitude objects. This distinction still held when adjusting for the relationship between morality and negativity. Further, real websites known by the United States government to be organized hate groups used significantly more moral language in expressing their beliefs as compared with users on complaint forums venting their dislike. Of note, we found an order effect in Study 1 such that differences between hate and dislike were less evident when participants were asked to generate disliked objects first. This suggests that people spontaneously think about objects that are closer to objects they extremely dislike or hate when asked about dislike without an explicit reference to hate.

Regarding the intensity hypothesis, we found mixed evidence for the role of negativity in distinguishing hate expressions from dislike. In Studies 1 and 2, hated attitude objects were rated as more negative than disliked attitude objects, even after controlling for morality, suggesting that both morality and negativity independently contribute to hate. These results are aligned with recent work by Martinez et al. (2022), who found increased ratings in 11 self-reported negative emotions in response to hated compared to disliked targets. However, these authors do not explore differences between hated and extremely disliked objects. We find this to be a relevant comparison in the light of Study 3, where we find that negativity differences between hated and extremely disliked objects vanished after controlling for morality, suggesting that differences in morality accounted for observed differences in negativity. Further, in Study 4, online expressions of hate did not use more negative language than online expressions of dislike. Thus, whereas hate and dislike seem to differ in both intensity and morality, it is possible that hate and extreme dislike differ mainly in the morality dimension. Future studies would benefit from employing scales and statistical techniques that allow researchers to obtain uncorrelated measures of negativity, attitude strength, and morality to better assess the independent contribution of each of these constructs in distinguishing hate from dislike.

Although it seems easy to recognize expressions of hatred when we see it—at Nazi rallies or ethno-cultural genocide—it is still poorly understood from a scientific perspective. Our studies find that morality is a key ingredient that differentiates the conceptualization of hate from dislike in the minds of many people. These studies offer a springboard for empirical research into the psychology of hate. Centuries of philosophical theory have laid the groundwork for more rigorous empirical investigation. For example, our review of the literature raised the possibility that hate is motivational. Rempel & Burris (2005) suggested that we will ignore disliked objects but will wish to harm hated ones. In line with this, people feel more inclined to engage in attack-oriented behaviors when they experience hate versus dislike (Martínez et al., 2022). Further, while the present laboratory studies manipulated the type of attitude object generated to test differences between groups, further research could reverse the relationship between our independent and dependent variables. An important test of the connection between hate and morality would be to determine if experimentally inducing moral emotions could create hate in a laboratory setting. However, the ethics of doing so must be carefully considered.

One potential alternative explanation is that our instructions to generate hated versus disliked attitude objects elicited different classes of attitude objects (e.g., people and groups versus concepts and beliefs). However, when we explicitly instructed people to generate different classes of attitude objects, we found that the difference between hate versus dislike was robust across these classes. It is also possible that hated versus disliked attitude objects differed systematically in level of abstraction. Some work has found that people more readily apply their moral principles to the psychologically distant (Eyal et al., 2008). Perhaps hated attitude objects are more psychologically distant or higher in abstraction? Another alternative explanation for our findings is that disliked versus hated objects do not need to have an actual antecedent: whereas people may not know why they dislike something, hatred may be more readily associated with a specific experience, making it easier to link to morality. Future research should address these possibilities.

An additional reason we believe the differences between hate and dislike extend beyond these issues is our study of online hate groups. The websites we explored did not require users to list attitudes objects they hated. In fact, many online hate groups actively disavow their categorization as "hate groups" and the content of their websites often focused on their core values (e.g., "*…teaches lessons of morality and nobility, to walk as a proud White individual in a world where being White is now considered wrong"*). Their websites were identified as hate groups by third parties. Our analyses nevertheless found much higher expressions of morality on these hate websites as compared to complaint forums, both about objects and corporate groups. Together with our lab experiments, this gives us confidence that the difference between hate and dislike goes beyond simple semantics.

Hate as Emotion

The current research relied on self-reports and content coding which provides a modest scope for understanding the rich affective experience of hate. At present, it is impossible to determine if hate causes a feeling state or if labeling an experience as hate is a consequence of an emotional experience (or both). Importantly, our use of the term hate does not imply that it is a basic emotion. Our belief is that the psychological state we colloquially associate with hatred is actively constructed like other complex emotional states rather than a natural kind (see Barrett, 2006). While this is beyond the scope of the present work, these are important distinctions that should be examined in future research on the psychology of hate.

On a related note, while the conceptualization of anger, contempt, and disgust as distinctively moral emotions continues to receive empirical support (see for instance Steiger &

Reyna, 2017), other scholars have challenged this view arguing that disgust may have a broader role beyond morality or that anger can be triggered by other moral transgression beyond autonomy (see Lomas, 2019). Thus, our results on the differences in moral emotions between hated and disliked attitude objects should be treated with caution: whereas these emotions may be necessary for hatred to arise, they may not be sufficient. Whereas higher ratings in anger, contempt, and disgust were to be expected in the hatred versus dislike condition, they should not be taken by themselves as unequivocal proof of the association between hatred and morality.

The results of the present research might, eventually, be fruitfully applied to psychological or behavior interventions against hate. For instance, work on relations between Israelis and Palestinians suggests that hatred toward the out-group differs from anger in terms of profiling the out-group as evil and intentionally causing harm (Halperin, 2008; see also Parker & Janoff-Bulman, 2013). Yet such conflict may be ameliorated and peace proposals more likely to be adopted when the out-group is willing to compromise sacred values—rather than economic concessions (Ginges et al., 2007). Thus, acknowledging and leveraging the moral concerns associated with hatred may provide an important avenue for addressing intergroup (as well as interpersonal) conflict. We urge research in these areas to continue this line of inquiry in the hopes of designing and testing interventions to alleviate social conflict.

Finally, we note that the samples of our first three studies were undergraduate students from Canada and the US. This poses a limitation in terms of the generalizability of the findings of these studies, which have been drawn from western educated individuals from industrialized, rich, democratic societies (WEIRD, see Henrich et al., 2010). We attempted to overcome this limitation in Study 4, where we obtained samples from a more ecological environment (websites with English-speaking audiences). Because different cultures could have different conceptualizations of hate and dislike, future research should further address this constraint by including cross-national representative samples.

Conclusion

Hate is undeniably a topic of considerable theoretical and practical interest. One only need open the newspaper or turn on the television to encounter daily instances of hate from rallies in Charlottesville, Virginia to terrorist attacks in Niger. The rise in hate crimes in the US suggests this issue is as urgent as ever. Although hate is easily recognized, the empirical literature has largely ignored the topic. Our finding that morality plays a role in differentiating lay theories of hate from dislike helps to address the paucity of research on this important topic. However, much more work needs to be done before psychologists will have the capacity to address this issue in a meaningful way.

References

Allport, G., Clark, K., & Pettigrew, T. (1954). *The nature of prejudice*. Addison-Wesley Pub. Co.

Associated Press (2013). Employees rate their employers, CEOs on Glassdoor. *Canadian Broadcasting Company.* Retrieved from https://www.cbc.ca/news/business/employees-rate-their-employers-ceos-on-glassdoor-1.1314945

Atran, S., Sheikh, H., & Gomez, A. (2014). Devoted actors sacrifice for close comrades and sacred cause. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(50), 17702–17703. https://doi.org/10.1073/pnas.1420474111

Aumer-Ryan, K., & Hatfield, E. (2007). The designs of everyday hate: A qualitative and quantitative analysis. *Interpersona*, *1*, 143–172. https://doi.org/10.5964/ijpr.v1i2.11

Barrett, L. F. (2006). Are emotions natural kinds? *Perspectives in Psychological Science, 1*, 28–58. https://doi.org/10.1111/j.1745-6916.2006.00003.x

Bartlett, S. (2005). *The pathology of man: A study of human evil*. Charles C Thomas Publisher.

Ben-Ze'ev. (2000). *The subtlety of emotions*. The MIT press.

Berman, R. (2017). Hate crimes in the United States increased last year, the FBI says. *The Washington Post.*

Brown, A. (2017). What is hate speech? Part 1: The Myth of Hate. *Law and Philosophy*, *36*(4), 419–468. https://doi.org/10.1007/S10982-017-9297-1

Capraro, V., & Rand, D. G. (2017). Do the Right Thing: Preferences for Moral Behavior, Rather Than Equity or Efficiency per se, Drive Human Prosociality. *Judgment and Decision Making*, 13(1), 99-111. https://doi.org/10.2139/ssrn.2965067

Cavalli-Sforza, L. L., Feldman, M. W., Chen, K. H., & Dornbusch, S. M. (1982). Theory and observation in cultural transmission. *Science*, 218(4567), 19–27. https://doi.org/10.1126/science.7123211

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2002). Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. In *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd Edition). Routledge. https://doi.org/10.4324/9780203774441

de Freytas-Tamura, K. (2017). U.K. reports big rise in hate crime, citing Brexit and terrorist attacks. *The New York Times*.

de Houwer, J., & Hughes, S. (2020). Learning to Like or Dislike: Revealing Similarities and Differences Between Evaluative Learning Effects. *Current Directions in Psychological Science*, 29(5), 487-49. https://doi.org/10.1177/0963721420924752

Dozier, R. W. (2002). *Why we hate: understanding, curbing, and eliminating hate in ourselves and our world*. Contemporary Books.

Edwards, T. (1908). *A cyclopedia of laconic quotations*. F. B. Dickerson Co.

Eyal, T., Liberman, N., & Trope, Y. (2008). Judging near and distant virtue and vice. *Journal of Experimental Social Psychology*, *44*, 1204–1209. https://doi.org/10.1016/j.jesp.2008.03.012

Fehr, B., & Russell, J. A. (1984). Concept of emotion viewed from a prototype perspective. *Journal of Experimental Psychology*, *113*, 464–486. https://doi.org/10.1037/0096-3445.113.3.464

Fischer, A. H., & Roseman, I. J. (2007). Beat Them or Ban Them: The Characteristics and Social Functions of Anger and Contempt. *Journal of Personality and Social Psychology*, *93*(1), 103–115. https://doi.org/10.1037/0022-3514.93.1.103

Fischer, A., Halperin, E., Canetti, D., & Jasini, A. (2018). Why We Hate. *Emotion Review*, *10*(4), 309–320. https://doi.org/10.1177/1754073917751229

Fitness, J., & Fletcher, G. J. O. (1993). Love, Hate, Anger, and Jealousy in Close Relationships: A Prototype and Cognitive Appraisal Analysis. *Journal of Personality and Social Psychology*, *65*(5), 942–958. https://doi.org/10.1037/0022-3514.65.5.942

Fitzgerald, L. F., Swan, S., & Magley, V. J. (1997). But was it really sexual harassment?: Legal, behavioral, and psychological definitions of the workplace victimization of women. In W. O'Donohue (Ed.), *Sexual harassment: Theory, research, and treatment* (pp. 5–28). Allyn & Bacon.

Frijda, N., Mesquita, B., Sonnemans, J., & van Goozen, S. (1991). The duration of affective phenomena or emotions, sentiments and passions. In *International Review of Studies on Emotion* (pp. 187–225). Wiley.

Ginges, J., Atran, S., Medin, D., & Shikaki, K. (2007). Sacred bounds on rational resolution of violent political conflict. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(18), 7357–7360. https://doi.org/10.1073/pnas.0701768104

Glaser, J., Dixit, J., & Green, D. (2002). Studying hate crime with the internet: What makes racists advocate racial violence? *Journal of Social Issues*, *1*, 177–193. https://doi.org/10.1111/1540-4560.00255

Green, D., Abelson, R., & Garnett, M. (1999). The distinctive political views of hate-crime perpetrators and White supremacists. In D. A. Prentice & D. T. Miller (Eds.), *Cultural divides: Understanding and overcoming group conflict* (pp. 429–464). Russel Sage Foundation.

Green, P., MacLeod, C., & Alday, P. (2016). *Package 'simr'.* [Computer software] https://cran. r-project. org/web/packages/simr/index. html.

Haidt, J., Rosenberg, E., & Hom, H. (2003). Differentiating diversities: Moral diversity is not like other kinds. *Journal of Applied Social Psychology*, *33*(1), 1–36. https://doi.org/10.1111/j.1559-1816.2003.tb02071.x

Halperin, E. (2008). Group-based hatred in intractable conflict in Israel. *Journal of Conflict Resolution*, *52*, 713–736. https://doi.org/10.1177/0022002708314665

Halperin, E. (2011). Intergroup hatred: Psychological dimensions. In D. J. Christie (Ed.), *The Encyclopedia of Peace Psychology* (pp. 557–561). Wiley-Blackwell. https://doi.org/10.1002/9780470672532.wbepp138

Halperin, E., Canetti, D., & Kimhi, S. (2012). In Love With Hatred: Rethinking the Role Hatred Plays in Shaping Political Behavior. *Journal of Applied Social Psychology*, *42*(9), 2231–2256. https://doi.org/10.1111/j.1559-1816.2012.00938.x

Henrich, J. (2022). Selective cultural processes generate adaptive heuristics. *Science*, *376*(6588), 31–32. https://doi.org/10.1126/science.abo0713

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Holpuch, A. (2014). Almost 100 hate-crime murders linked to single website, report finds. *The Guardian*.

Hoover, J., Atari, M., Mostafazadeh Davani, A., Kennedy, B., Portillo-Wightman, G., Yeh, L., & Dehghani, M. (2021). Investigating the role of group-based morality in extreme behavioral expressions of prejudice. *Nature Communications*, *12*(1), 1–13. https://doi.org/10.1038/s41467-021-24786-2

Hox, J. (1998). Multilevel Modeling: When and Why. In: Balderjahn, I., Mathar, R., Schader, M. (eds) Classification, Data Analysis, and Data Highways. Studies in Classification, Data Analysis, and Knowledge Organization. Springer. https://doi.org/10.1007/978-3-642-72087-1_17

Krug, E. G., Dahlberg, L. L., Mercy, J. A., Zwi, A. B., & Lozano, R. (2002). World report on violence and health. *Lancet*, 360(9339), 1083-8. https://doi.org/10.1016/S0140-6736(02)11133-0

Lomas, T. (2019). Anger as a moral emotion: A "bird's eye" systematic review. *Counselling Psychology Quarterly*, *32*(3–4), 341–395. https://doi.org/10.1080/09515070.2019.1589421

Martínez, C. A., van Prooijen, J. W., & Lange, P. A. M. V. (2022). Hate: Toward Understanding Its Distinctive Features Across Interpersonal and Intergroup Targets. *Emotion*, *22*(1), 46–63. https://doi.org/10.1037/EMO0001056

Meier, J. S. (1993). Notes from the Underground: Integrating Psychological and Legal Perspectives on Domestic Violence in Theory and Practice. *Hofstra Law Review*, *21(4)*, 1295-1366. http://scholarlycommons.law.hofstra.edu/hlrAvailableat:http://scholarlycommons.law.hofstra.edu/hlr/vol21/iss4/4

Ortony, A., Clore, G., & Collins, A. (1990). *The cognitive structure of emotions*. Cambridge University Press. https://doi.org/10.1017/CBO9780511571299

Parker, M. T., & Janoff-Bulman, R. (2013). Lessons from morality-based social identity: The power of outgroup "hate," not just ingroup "love." *Social Justice Research*, *26*(1), 81–96. https://doi.org/10.1007/s11211-012-0175-6

Rempel, J. K., & Burris, C. T. (2005). Let me count the ways: An integrative theory of love and hate. *Personal Relationships*, *12*(2), 297–313. https://doi.org/10.1111/j.1350-4126.2005.00116.x

Royzman, E. B., McCauley, C., & Rozin, P. (2006). From Plato to Putnam: Four Ways to Think About Hate. In *The Psychology of Hate.* (pp. 3–35). American Psychological Association. https://doi.org/10.1037/10930-001

Rozin, P. (1999). The Process of Moralization. *Psychological Science*, *10*(3), 218–221. https://doi.org/10.1111/1467-9280.00139

Rozin, P., & Fallon, A. E. (1987). A Perspective on Disgust. *Psychological Review*, *94*(1), 23–41. https://doi.org/10.1037/0033-295X.94.1.23

Skitka, L. J. (2002). Do the means always justify the ends or do the ends sometimes justify the means? A value protection model of justice reasoning. *Personality and Social Psychology Bulletin*, *28*, 588–559. https://doi.org/10.1177/0146167202288003

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

Skitka, L. J., & Mullen, E. (2002). The dark side of moral conviction. *Analyses of Social Issues and Public Policy*, *2(1)*, 35–41. https://doi.org/10.1111/j.1530-2415.2002.00024.x

SPLC. (2017). Active Hate Groups 2016. In *Intelligence Reports*. Southern Poverty Law Center.

Staub, E. (2004). The origins and evolution of hate, with notes on prevention. In *Psychology of Hate* (pp. 51–65). American Psychological Association. https://doi.org/10.1037/10930-003

Steiger, R. L., & Reyna, C. (2017). Trait contempt, anger, disgust, and moral foundation values. *Personality and Individual Differences*, *113*, 125–135. https://doi.org/10.1016/j.paid.2017.02.071

Sternberg, R. J. (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

Sternberg, R. J. (2005). *The Psychology of Hate.* (R. J. Sternberg, Ed.). American Psychological Association. https://doi.org/10.1037/10930-000

Van Bavel, J., Packer, D., Haas, I., & Cunningham, W. (2012). The Importance of Moral Construal: Moral versus Non-Moral Construal Elicits Faster, More Extreme, Universal Evaluations of the Same Actions. *PLoS ONE*, *7*(11). https://doi.org/10.1371/journal.pone.0048693

Watson, D., & Clark, L. A. (1994). *The PANAS-X: Manual for the positive and negative affect schedule-expanded form*. The University of Iowa. https://doi.org/10.17077/48vt-m4t2

White, R. K. (1996). Why the Serbs fought: Motives and misperceptions. *Peace and Conflict: Journal of Peace Psychology*, *2*(2), 109–128. https://doi.org/10.1207/s15327949pac0202_2

Figure 1: "Wordles" illustrating the frequency with which participants listed each attitude object in Study 1 (Top: Dislike, Bottom: Hate). The larger the word appears, the more frequently it was listed.

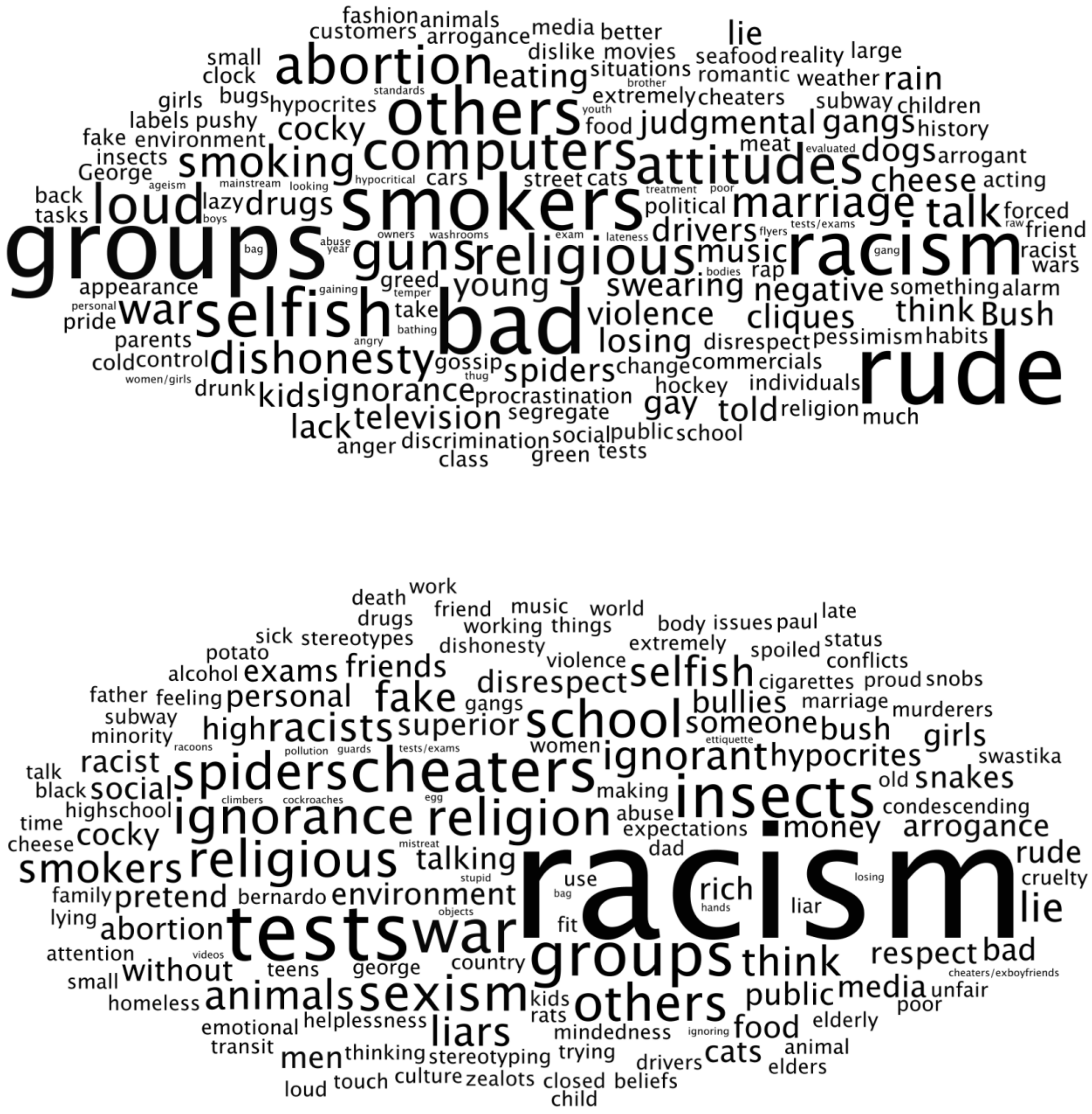
fashion animals media better lie
customers arrogance dislike movies seafood reality large
small abortion eating situations romantic weather rain
clock standards brother extremely cheaters subway children
girls bugs hypocrites others youth food judgmental gangs history
labels pushy cocky computers meat evaluated dogs arrogant
fake environment insects smoking attitudes cheese acting
George mainstream hypocritical cars street cats treatment poor
back ageism looking loud lazy drugs smokers political marriage talk forced
tasks boys abuse owners washrooms exam lateness tests/exams raw friend
groups bag year guns religious drivers flyers racism gang racist
music rap bodies wars
appearance gaining greed young swearing negative something alarm
personal temper war selfish bad violence cliques think Bush
pride take bathing losing disrespect pessimism habits
parents angry dishonesty gossip change commercials
cold control thug spiders hockey individuals rude
women/girls drunk kids ignorance procrastination gay told religion much
lack television segregate social public school
anger discrimination class green tests
death work
drugs friend music world late
sick stereotypes working things body issues paul
potato dishonesty extremely spoiled status
alcohol exams friends violence conflicts
father feeling personal fake disrespect selfish cigarettes proud snobs
subway gangs bullies marriage murderers
minority high racists superior school someone bush girls
racoons pollution guards tests/exams women ettiquette
racist spiders cheaters ignorant hypocrites swastika
talk social making old snakes
black climbers cockroaches egg insects condescending
highschool ignorance religion abuse
time cocky expectations money arrogance
cheese mistreat dad rude
smokers religious talking racism cruelty
family stupid use losing
pretend bernardo environment bag rich lie
lying objects hands liar
abortion tests war fit
attention videos groups think respect bad
without teens george country cheaters/exboyfriends
small animals sexism kids others public media unfair
homeless rats poor
emotional helplessness liars food elderly
transit mindedness ignoring
men thinking trying cats animal
stereotyping drivers elders
loud touch culture zealots closed beliefs
child

Figure 2: The effect of hate versus dislike on morality judgments **(a)** and moral emotions **(b)** in Study 1. Hated objects were rated as significantly more tied to core moral beliefs and moral emotions than disliked attitude objects. Disliked attitude objects, rated after hated attitude objects, received attenuated morality ratings.

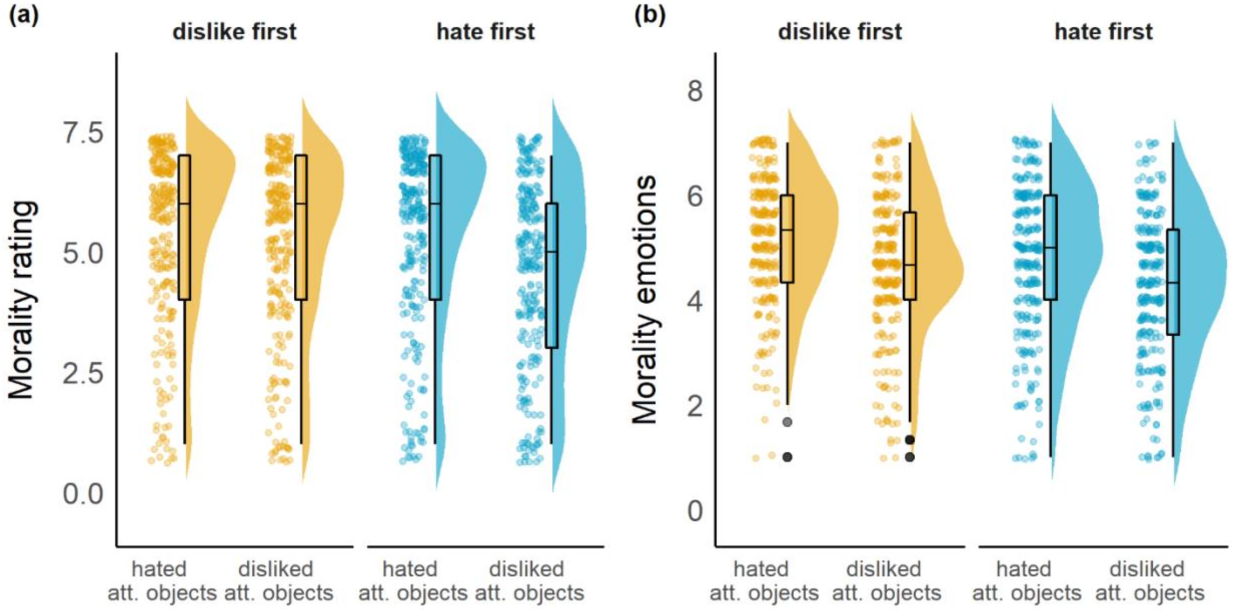

Figure 3: The effect of hate versus dislike on morality judgments **(a)** and moral emotions **(b)** in Study 2. Hated objects were rated as significantly more tied to core moral beliefs and moral emotions than disliked attitude objects.

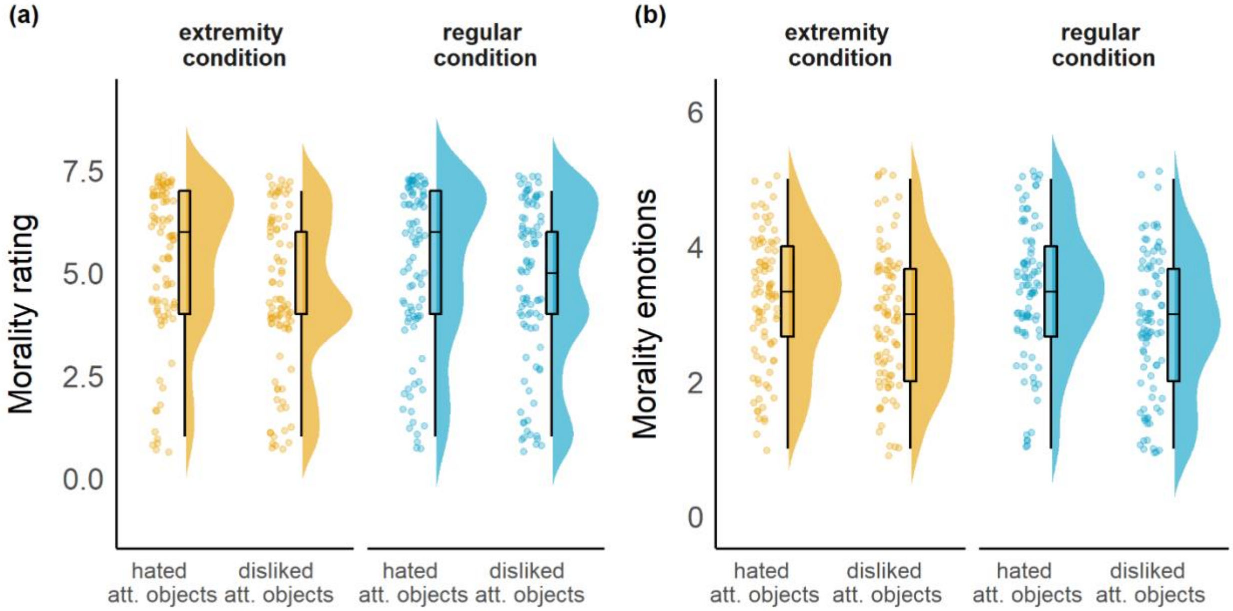

Figure 4. The effect of hate versus dislike on ratings of (a) negativity, (b) morality and (c) attitude strength in Study 3. Hated objects were rated as significantly more negative and were more associated with attitude strength than disliked objects but received similar ratings in negativity and attitude strength as extremely disliked objects. Hated objects were rated as more related to morality than extremely disliked attitude objects and disliked attitude objects.

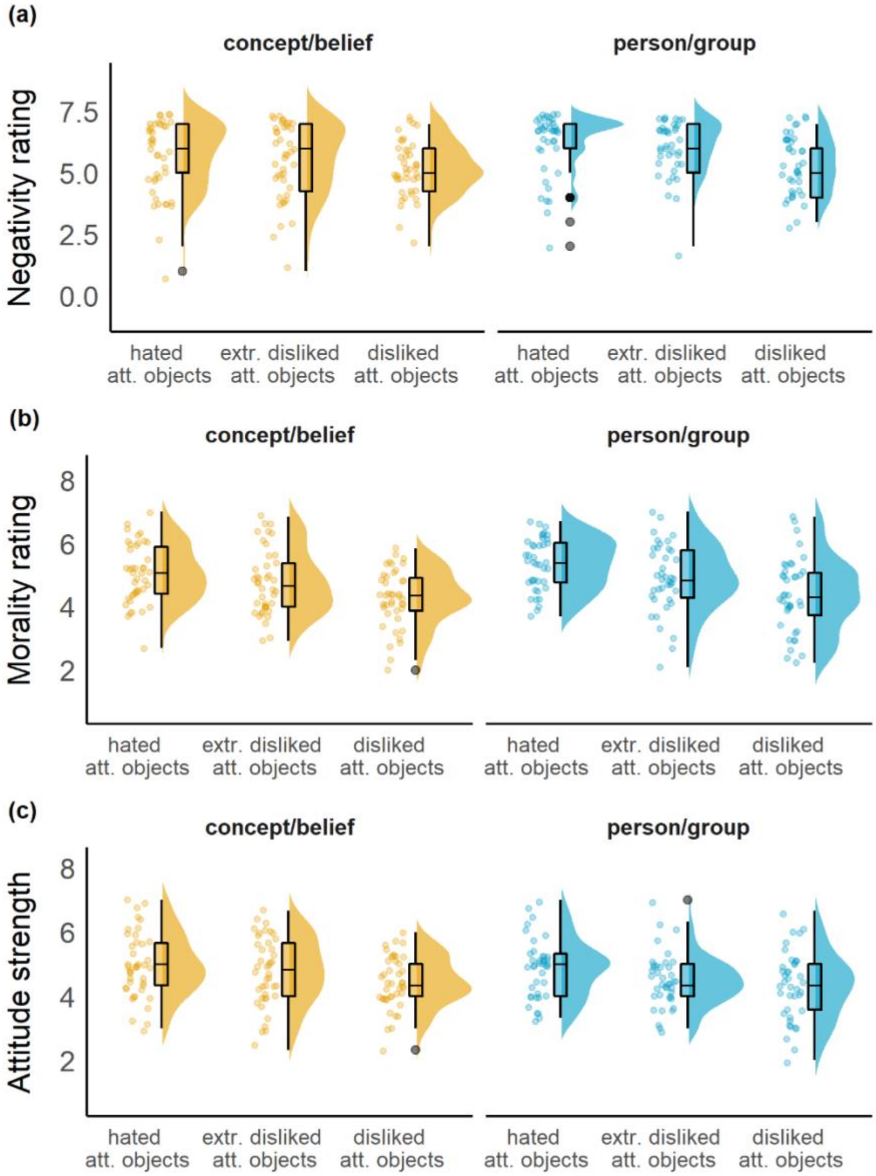

Table 1. Bivariate correlations of negativity, morality, and moral emotions (Studies 1 & 2). Bivariate correlations of negativity, morality, and attitude strength (Study 3). For ease of interpretation, we show the zero-order correlations of the dependent variables treating each rating made by participants as independent. **p* <.05.

Study 1

	1	2	3
1. Morality	-		
2. Negativity	.20*	-	
3. Moral Emotions	.42*	.30*	-

Study 2

	1	2	3
1. Morality	-		
2. Negativity	.16	-	
3. Moral Emotions	.29*	.23*	-

Study 3

	1	2	3
1. Morality	-		
2. Negativity	.59*	-	
3. Attitude strength	.57*	.42*	-

Supplementary materials

Supplementary Table 1. Descriptives of each of the assessed moral emotions in Study 1 and 2.

	Overall		**Hated att. objects**		**Disliked att. objects**	
	Study 1 *M (SD)*	Study 2 *M (SD)*	Study 1 *M (SD)*	Study 2 *M (SD)*	Study 1 *M (SD)*	Study 2 *M (SD)*
Anger	4.95 (1.84)	3.19 (1.11)	5.27 (1.80)	3.60 (1.36)	4.63 (1.82)	3.25 (1.38)
Contempt	4.10 (1.70)	2.09 (1.09)	4.17 (1.78)	2.32 (1.35)	4.03 (1.61)	2.18 (1.24)
Disgust	5.05 (1.75)	3.30 (1.17)	5.37 (1.68)	3.84 (1.84)	4.73 (1.77)	3.19 (1.47)