

8-2019

## Hating Evil: Understanding the Role of Evil in Interpersonal Hate

Carmen Merrick  
*University of Arkansas, Fayetteville*

Follow this and additional works at: <https://scholarworks.uark.edu/etd>



Part of the [Cognition and Perception Commons](#), [Experimental Analysis of Behavior Commons](#), and the [Social Psychology Commons](#)

---

### Recommended Citation

Merrick, Carmen, "Hating Evil: Understanding the Role of Evil in Interpersonal Hate" (2019). *Theses and Dissertations*. 3380.

<https://scholarworks.uark.edu/etd/3380>

This Dissertation is brought to you for free and open access by ScholarWorks@UARK. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of ScholarWorks@UARK. For more information, please contact [ccmiddle@uark.edu](mailto:ccmiddle@uark.edu).

Hating Evil: Understanding the Role of Evil in Interpersonal Hate

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy in Psychology

by

Carmen Merrick  
Ouachita Baptist University  
Bachelor of Science in Psychology, 2011  
University of Texas at Dallas  
Master of Science in Applied Cognitive Neuroscience, 2014  
University of Arkansas  
Master of Arts in Psychology, 2017

August 2019  
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

---

Denise Beike, Ph.D.  
Dissertation Director

---

Scott Eidelman, Ph.D.  
Committee Member

---

Jennifer Veilleux, Ph.D.  
Committee Member

## Abstract

Research has demonstrated that hate is a protective emotional response to perceived evil, yet the dimensions by which people perceive evil have not been clearly identified. Research has also indicated that it is evil to feel hate, which presents an interesting paradox: if hate protects us from evil, then how can it be evil to feel hate? The present research attempts to identify the dimensions of evil and elucidate the relationship between hate and evil by comparing it to the relationship between dislike and evil. Study 1 tested how participants identified evil in third person scenarios. As predicted, evil was labeled most often in scenarios that depicted intentional acts on the part of the target, though contrary to predictions evil was not labeled as often in scenarios that depicted responsible acts on the part of the target. Study 2 relied on participants' own experiences of dislike and hate, or the experiences of someone, they knew toward either a group or individual. As predicted, hate and dislike differed in intensity and kind, though less so in kind than anticipated. Also as predicted, differentiations emerged within hate that suggested that hate toward groups, specifically when experienced by others, is considered more evil and less justified than hate experienced by the self. The consistency of these results with extant theories of hate and evil are discussed as well as potential explanations for the differential relationship of hate and evil.

## Table of Contents

I.	Introduction.....	1
	A. The Problem with Hate.....	2
	B. The Dislike Problem: Differentiating Hate and Dislike.....	4
	C. The Anger Problem: Differentiating Hate and Anger.....	5
	D. The Evil Problem: What is Evil?.....	8
	E. The Two Hates Problem.....	14
	F. Intergroup vs. Interpersonal Hate: A Morality Issue?.....	17
	G. Statement of the Problem.....	21
II.	Study 1.....	25
	A. Method.....	25
	B. Data Analysis Plan.....	30
	C. Results and Discussion.....	32
III.	Study 2.....	37
	A. Method.....	37
	B. Data Analysis Plan.....	41
	C. Results and Discussion.....	44
IV.	General Discussion.....	52
	A. Limitations.....	57
	B. Future Directions.....	58
	C. Conclusions.....	58
V.	References.....	60
VI.	Tables.....	67

VII. Appendices .....	78
-----------------------	----

## Introduction

Martin Heidegger was one of the foremost thinkers of the twentieth century. He contributed to many schools of thought and provided a foundation for many thinkers who came after him. He was also a zealous Nazi. An examination of Heidegger's life and philosophy has led to a decades-long debate about whether or not his ideas should still hold influence. One author has called for the entire canon of Heidegger's work to be treated as hate speech (Faye, 2009). He argues that Heidegger's Nazi beliefs, which were so full of vitriol and hate, make him evil and his influence should therefore be erased, regardless of the positive consequences it might have had on others. It begins to feel as if the author would erase Heidegger altogether if he could. Few would argue that Heidegger's hate was anything but evil. Yet research has demonstrated that hate protects us from others we consider evil (Merrick & Beike, 2019). How can feeling hate be evil and at the same time protect us from the evil of others?

The relationship between hate and evil is an exceedingly complex one. The words are often used interchangeably and hate and evil are considered to be the same or rooted in the same causes (Baumeister 1996; Sternberg, 2003). In our own research, we have demonstrated that hate labels what is evil and protects the self from it (Merrick & Beike, 2019), but other researchers have demonstrated people avoid feeling hate because they believe it incites evil in the self (Halperin, 2008). If hate protects us from evil, how can it be evil to feel hate? This question is approached as a scientific inquiry, but it is just as important to understand people's perceptions of this question and of the concepts of hate and evil (James, 1890; Johnson-Laird & Oatley, 1992). It is not only important to empirically explore the relationship between hate and evil, but to understand people's *beliefs* about hate and evil and how these concepts are perceived to function and interact.

## **The Problems with Hate**

Broadly speaking, there are two main problems with the study of hate. The first is that hate is an undeniably difficult construct to define. One reason for this difficulty is the disagreement surrounding what kind of a construct hate actually is (see Royzman, McCauley, and Rozin, 2003 for review). Hate has been described as a discrete emotion (Elster, 1999), separable and distinguishable from other negative emotions. It has also been described as a complex emotion that is comprised of other emotions, including a combination of anger and fear (McCauley, 2002), and a combination of anger, contempt, and disgust (Matsumoto, Hwang, & Frank, 2017; Sternberg, 2003). Relatedly, hate has been described as a perversion of other negative emotions including disgust (Oatley & Johnson-Laird, 1987) and anger (Dozier, 2002; Frijda, 1986). Finally, hate has been described as a sentiment, or a negative identification that informs emotions but is separate from them (Gervais & Fessler, 2017; Shand, 1920), and an emotional attitude akin to extreme dislike (Ben Ze'ev, 1992; Ekman, 1992). As Royzman, McCauley, and Rozin (2003) point out, there is no generally accepted definition of hate.

In a study in which participants were asked to describe the nature of hate, hate was typically described as an emotion (Aumer-Ryan & Hatfield, 2007). Similarly, in a study in which participants provided characteristics of hate, we found that they described it as being brief, frequent, and intense (Merrick & Beike, 2019). In other words, participants characterized hate as emotion-like, rather than as a sentiment or emotional attitude. Both of the studies described above also demonstrated that hate tends to be associated with other negative emotions, though differently. Aumer-Ryan and Hatfield (2007) found that participants described hate as an extreme form of other negative emotions including dislike, anger, and disgust. In our study, participants reported that their hate co-occurs with other negative emotions including anger and

dislike. Notably, these studies differed in their procedure of asking participants about hate. The former utilized a qualitative method by having participants describe hate, while the latter specifically asked participants which emotions they experienced during an episode in which they experienced hate, utilizing a more quantitative methodology. Therefore, it is hard to determine whether hate merely co-occurs with other negative emotions or whether it consists of other negative emotions. Importantly, with the exception of Merrick and Beike (2019), the theories and studies described above do not differentiate between intergroup and interpersonal hate, which might account for the discord in the theoretical construct attributed to hate. For the purpose of the current study, and based on past research, hate is defined as *a negative emotional response to a perception that a target is evil or has acted immorally*.

The second broad issue with the study of hate is that, while there are several theoretical perspectives explicating it, there has been very little empirical exploration. Within the research that does exist on hate, there are two important observations to be made. First, hate is often compared with other negative emotions, namely anger and dislike. This body of research provides evidence that hate is not just an extreme or perverse manifestation of one of these emotions. The comparison of hate to each of these emotions elucidates different characteristics of hate. This literature and the subsequent conclusions about hate are discussed in the following sections. The second observation about the research on hate is that it is inconsistent in regard to its target. Some research focuses on intergroup hate, or hate toward a person or group of people based on group membership, while other research focuses on interpersonal hate, or hate toward an individual based on character or actions unrelated to group membership. The possible importance of this inconsistency and the subsequent conclusions about hate are also discussed.

## **The Dislike Problem: Differentiating Hate and Dislike**

In determining whether there is a difference in hate and dislike, two important dimensions have been identified: difference in degree and difference in kind (Van Bavel, Ray, Granot, & Cunningham, 2018). According to Darwin, dislike and hate exist along the same spectrum, separated only by intensity, and therefore are only different by degree (Darwin, 1872). A similar perspective has been offered more recently, in which hate is argued to be a global negative attitude, similar to dislike (Ben-Ze'ev, 1992; Ben Ze'ev, 2000). This theoretical assumption has gone so far as to infiltrate the scientific study of hate. In one study in which participants were asked to interpret characters' emotional reactions in scenarios, hate and dislike were part of the same item and treated synonymously (Fitness & Fletcher, 1993).

A difference of kind suggests that hate and dislike are distinct negative affective states, separated by more than just degree of intensity. One suggestion for the difference in hate and dislike is the nature of their motivation. Allport (1954) suggested that, similar to dislike, hate is a sentiment, but that it is motivated by ethical concerns. In a series of studies, the question of whether hate and dislike differ in degree or in kind was addressed by measuring the negativity (degree) and morality (kind) of both targets of hate and dislike (Van Bavel et al., 2018). Results revealed that hate and dislike differ in both degree and in kind. Specifically, hate is more intense and more connected to moral convictions than dislike. Furthermore, when controlling for negativity, hated targets are still considered more connected to morals. These results support Allport's (1954) claim that, in order to experience hate, a person must consider the target of their hate to have violated something morally valuable to them. To my knowledge, Van Bavel and colleagues' manuscript is the first empirical comparison of hate and dislike. This study successfully differentiates hate from dislike and indicates that hate is concerned with some moral

violation. However, though hate was determined to be more concerned with morals and moral convictions, the nature of that moral connection was not explored.

Another explanation for the differences in hate and dislike, in both degree and in kind, is the construct that each represents. As previously mentioned, research has demonstrated that hate is experienced as an emotion. There is theoretical evidence that dislike, too, is an emotion (Roseman, 1984), though, most research indicates that dislike is an attitude (Collins, 1970; Petty, Ostrom, & Brock, 2014; Zimbardo, Ebbesen, & Maslach, 1977). Attitudes are more permanent and less intense toward their object, whereas emotions are brief, intense, and mobilizing (Gervais & Fessler, 2017; Roseman, 1984). According to one theory, sentiments, or functional networks of affective constructs, link emotions and attitudes in regard to a common object (Gervais & Fessler, 2017). For instance, a person might dislike broccoli, which is considered an attitude. However, the emotions the person has related to broccoli are likely situational – she might feel disgust when broccoli is presented to her, or anger toward a person who tries to serve her broccoli. The emotions and attitudes she has that relate to broccoli exist within a network that pertain specifically to that object. If hate is an emotion and dislike is an attitude, then it is likely hate and dislike differ in degree, as emotions are more intense than attitudes. Hate and dislike also likely differ in kind, since they are different constructs, and therefore would be differentially related to the moral convictions of the person experiencing them. In the present research, the relationship between hate and dislike will be examined more thoroughly.

### **The Anger Problem: Differentiating Hate and Anger**

There are two characteristics that are important to address in distinguishing hate and anger. The first is the evaluative appraisal that serves as an impetus for hate compared with anger. Appraisal theories of emotion assert that emotions are elicited and differentiated from one

another based on a person's subjective evaluation of a situation (Roseman, 1984; Scherer, 1984). Appraisals are sufficient, but not necessary to cause emotions (Ellsworth, 2013; Frijda, 1993). Theoretical evidence, which often compares hate and anger, suggests that hate is the emotional response to evaluations about the character of the target, whereas anger is the emotional response to the target's actions (Allport, 1954; Ben Ze'ev, 1992; Elster, 1999; Fischer et al., 2018; Halperin, 2008; Horowitz, 2001). In other words, "hatred says that 'he is bad,' anger that 'he did something bad to me'" (Elster, 1999, p. 65).

In an experimental examination of interpersonal hate compared with anger, when participants experienced hate, it was characterized by an attribution that the target and his or her actions were evil (Merrick & Beike, 2019). In contrast, while the target of anger may have committed some harm against the participant, the act and the target were not considered evil. The attribution of evil associated with hate has also been demonstrated in the examination of intergroup hate compared with anger (Halperin, 2008). Both hate and anger were associated with appraisals that the outgroup had committed an unjust action, but only hate was associated with an appraisal that the target was evil. The implication of a character judgment attached to hate but not anger suggests that hate serves a purpose that is discrete from that of anger, despite theories that suggest that hate is a perversion of anger.

Another important characteristic that is used to compare hate and anger is the action tendency and the motivation for that action tendency. According to appraisal theories, action tendencies are responsive actions motivated by emotions (Frijda, 1986; Moors, Ellsworth, Scherer, & Frijda, 2013). Though these actions are not always (or even usually) realized, they increase the person's tendency to engage in goal-related actions associated with the emotion-eliciting event. In general, the action tendency of hate has been described as one of destruction,

elimination, or aggression (Elster, 1999; Fischer et al., 2018; Halperin, 2008; Sternberg, 2003;). However, other theorists have suggested that the associated action tendency of hate is one of avoidance (Beck, 1999; Fitness & Fletcher, 1993).

In our own research, we explored the action tendency of interpersonal hate and found that it was consistently associated with avoidance across five studies, as opposed to anger which was characterized by a mixed action tendency of approach and avoidance (Merrick & Beike, 2019). In the final study of that set, we also examined the motivation for the action tendency associated with hate and anger. For hate, along with the anticipated action tendency of avoidance, there was an indication that the avoidance was driven by a desire to protect the self. For anger, the motivation of the action tendency was once again mixed. In other words, the experience of hate indicates to its host that the target is dangerous, harmful, or destructive. Anger tends to be a more flexible emotion that can have either an approach or avoidance action tendency and have a more general use. This motivation is further expounded upon by an earlier study, in which interpersonal hate and anger were explored in the workplace (Fitness, 2000). When participants reported that they experienced hate during anger-eliciting events, the presence of hate was negatively related to the perception that the incident had been resolved. In another study exploring interpersonal hate, Fitness and Fletcher (1993) found that, when feeling anger toward a spouse, participants tended to approach but when they felt hate they tended to avoid the spouse. These findings suggest that anger is constructive and motivates its subject to preserve or mend the relationship. Hate is destructive and motivated to avoid or eliminate the relationship. This was also found in a comparison of intergroup hate and anger (Halperin, 2008). Though the action tendency of intergroup hate was more mixed than we found in an examination of interpersonal hate, the motivation for the action tendency of intergroup hate was still associated with the

perception that there was no opportunity to change the relationship. Together, this research suggests that hate is concerned with the destruction of a relationship that has become harmful to the self.

Combining this comparison of the action tendency of hate and anger with the comparison of the appraisals that accompany hate, and anger and the comparison of hate and dislike, a more complete model takes form. Hate and dislike are differentiated, not just by intensity but by the intertwining of hate with some moral conviction (Van Bavel et al., 2018). In both interpersonal hate and anger, the target is perceived to have committed some personal harm or injustice toward the subject. However, in anger, though the offense may be exactly the same, the attribution differs. In anger, the offense does not lead to the complete destruction of the relationship. The target is perceived to have done something bad, but is not perceived to be a bad person, and so the motivation for any action tendency is to preserve the relationship. In interpersonal hate, the harm or injustice does not just hurt the subject, it is assumed to reveal something inherently evil about the target. Only a bad person could do what this person did. In this case, the motivation for the action tendency is to protect the self, even at the expense of the relationship. Though this is a profound revelation, it unfortunately does not completely illuminate hate. To do that, it is also necessary to understand evil.

### **The Evil Problem: What is Evil?**

The function of hate is to reveal to us what is evil and to help us avoid it. Arguably, hate either reveals to us the evil in others, the evil in ourselves, or both. What is clear is that hate and evil are inextricably linked. Hate reveals to us the evil in ourselves and others – but what is evil? Like hate, evil is an exceptionally fuzzy term that is often used without much consideration for its precise meaning. Like hate, there is no generally accepted definition of evil. For instance, in

defining hate, the target is often defined as evil without evil ever being expressly defined itself (Baumeister & Butz, 2005; Opatow, 2004). In theoretical work on hate, the presence of evil implies that a person is dangerous (Ben Ze'ev, 2000), or “intrinsically and irremediably bad” (Elster, 1999, p. 65). These characteristic terms do little to describe evil, though, as “bad” and “dangerous” are themselves fuzzy and in neither case are they quantifiably measured. A further complication arises from the application of the word “evil” to people, actions, motivations, and organizations (Singer, 2004). This general application of the word extends the difficulty that arises in trying to define it.

Besides being a generously applied term, evil has been explored across disciplines and perspectives, including social psychology, personality psychology, religious studies, and philosophy, which collectively provide different ideas about the nature of evil. For instance, having been born from a pursuit to understand the atrocities of World War II, social psychology has provided perhaps the most theories of evil. There are variations within them, but these theories all share a common theme: evil is not a disposition, it is a product of the situation or the influence of an organization. In other words, people are not evil; the situation makes people act in evil ways.

In contrast, philosophers and personality psychologists have suggested a dispositional perspective of evil. Dispositional perspectives assert that it is not the situation but the person's character that is evil. As discussed previously, it has been demonstrated that people attribute evil to the target of their hatred (Halperin, 2008; Merrick & Beike, 2019). Taking an extreme view of the dispositional nature of evil, Kant proposed that evil was not an individual difference but an innate proclivity of all people toward evil (Kant, 1785). Personality psychologists take a slightly less gloomy view, considering evil to be a combination of personality traits. Specifically, the

Dark Tetrad (previously the Dark Triad) is proposed as the basis of human evil (Book, Visser, & Volk, 2014; Book et al., 2016; Paulhus & Williams, 2002). The Dark Tetrad includes Machiavellianism (i.e., calculated manipulation), psychopathy (i.e., lack of empathy and increased predatory behavior), narcissism (i.e., grandiosity and excessive importance of the self), and sadism (i.e., enjoyment in harming others). Though the situationist and dispositional perspectives of evil differ, together they offer three main dimensions that people use to categorize others as evil including intentionality, responsibility, and immorality.

**Intentionality.** Both situationist and dispositional perspectives of evil indicate that one characteristic of evil is that it is intentional. Intentionality refers to the motivation of the target who commits evil. Put simply, the person has done something *on purpose*. According to a widely cited definition of evil from the situationist perspective, evil is “intentionally behaving – or causing others to act – in ways that demean, dehumanize, destroy, or kill innocent people” (Zimbardo, 2004, p. 4). In other words, though the person may be influenced by situational factors, their behavior is purposeful. For example, if one person accidentally kills another, then it is not considered evil. On the other hand, if a person intentionally murders another, then it is considered evil.

According to the dispositional model of evil, the person who is “evil” (someone who is high in the Dark Tetrad constellation of personality traits: narcissism, psychopathy, Machiavellianism, and sadism) intentionally causes harm to others, either out of pleasure or disregard for the welfare of others (Furham, Richards, & Paulhus, 2013). Both perspectives support the idea that, regardless of whether it is because the person is evil or because of the situation, the person is motivated to commit immoral acts.

**Responsibility.** A second dimension by which evil is considered through is the responsibility of the agent. Responsibility refers to whether or not the target made a rational choice in perpetrating the immoral acts, or in other words, the target should be held accountable for his or her actions. From the situationist perspective, the target must be aware of the consequences of their actions (Zimbardo, 2004). For instance, one suggestion is to consider the criteria by which we incarcerate those who have committed crimes (Darley, 1992). We consider whether the person was responsible for their actions by considering whether they could fathom the consequences for those actions and whether they had a choice in their actions. We also consider whether the person acted under duress. Referring to the previous example of intentional and unintentional murder, if a woman intentionally kills her husband, but the husband in question had systematically abused and terrorized the woman, then she may or may not be considered responsible for her actions, and therefore her actions might not be considered evil. This is not to say her actions are considered moral. Acts can be immoral without being evil. Evil entails both intentionality and responsibility.

The dimension of responsibility is also debated from the dispositional perspective. Moral development and the Dark Triad of traits have been linked to environmental rather genetic factors (Campbell, 2009). Additionally, a longitudinal study demonstrated that harsh and unpredictable environments in adolescence have downstream negative behavioral effects in early adulthood (Brumbach, Figueredo, & Ellis, 2010). The central question of this perspective is this: is the person still considered evil who has been nurtured into evil ways? For instance, a literary example of this question exists in the examination of Heathcliff in *Wuthering Heights* (Brontë, 1847). Initially, he is considered both responsible and intentional in his actions and the resulting evaluation leaves little doubt of his evil character. The reader then spends the rest of the novel

learning about Heathcliff's life story and the question of Heathcliff's responsibility for his evil character looms heavy throughout. Ultimately, it is not as important to discern what is or is not evil character, but to determine people's interpretations and evaluations of character. In other words, if people are not stopping to consider the backstory of the evil person, then they no doubt hold them responsible for their actions. Both responsibility and intentionality are hypothesized as necessary criteria for people's perceptions of evil. A person must choose to act and be motivated to act immorally to be considered evil.

Responsibility differs from intentionality in that a person might act intentionally but be in an altered state of mind or under duress. For example, if a woman is kidnapped and a bomb is forcibly strapped to her chest and she is told it will detonate and kill herself and others if she refuses to follow her kidnapers' instructions, she might intentionally follow their orders even if they caused harm to another person, but not be held responsible for her actions. Alternately, responsibility might also be present in the absence of intentionality. If a driver strikes and kills a pedestrian with his car but is texting and thus not paying sufficient attention to driving conditions, he may not have intentionally killed the pedestrian, but he would still be held responsible for his actions. Weiner (1993) argued that responsibility and intentionality are closely related but that responsibility is subject to judgments about moral justification whereas intentionality is not. In other words, an act can be intentional or unintentional, but the person who committed the act cannot be deemed responsible until the circumstances of the person are known. An additional difference, he argued, is that intentionality is a causal property and tells us about how the person carried out their actions, whereas responsibility is an inference we make about the person's character.

**Immorality.** In the study of evil, there has been some disagreement about the nature of the acts that must be perpetrated in order to be considered evil, but there is consensus in the assumption that some sort of immorality underlies evil. Moral Foundations Theory (MFT) argues that evil can exist based on a violation of any of the foundations of morality (Graham & Haidt, 2012). According to MFT, there are six moral foundations that make up people's moral systems based on which of the foundations they value (Graham, Haidt, & Nosek, 2009; Haidt & Joseph, 2004). According to MFT, evil is "anything that stands in the way of sacredness" (Graham & Haidt, 2012, p. 6). In contrast, the situationist perspective defines evil as intentional acts of personal harm, or lack of care, therefore violating the harm/care foundation (Baumeister, 1996; Staub, 1999; Zimbardo, 2004). What is clear from the research on evil is that it includes immorality. However, as argued earlier, an act can be immoral without being evil. It is hypothesized here that immorality is necessary but not sufficient for evil. Rather, it is immorality that is intentional and for which the agent is responsible.

**Evil Conclusions.** In the study of hate, the nature of evil is of secondary interest to people's beliefs about evil. If the same acts can instigate both anger and hate, but the determination of evil distinguishes one from the other, then it is important to understand how people define evil. It is not important that the target of hate is actually evil –this analysis is not intended as a way of determining what is truly good and truly evil. The central question of interest in the present research is, what do people believe about evil and what criteria they are using to make that determination? This is a distinction that is often made between scientific inquiry and folk psychology and it is particularly important to the current research (Johnson-Laird & Oatley, 1992). Though the two are often pitted against one another, this is unnecessary as one informs the other. Specifically, regarding evil, people are thought to have an intuition of

what is evil to help them avoid evil people (Darley, 1992). Using both the situationist and dispositional perspectives of evil, the dimensions that people use to evaluate evil are identified as intentionality, responsibility, and perception of immorality.

In the research on hate, one place where the problem of evil is particularly pronounced is in the differentiation of intergroup and interpersonal hate. The role of evil in these two constructs is explored and suggested as a possible component on which they differ. Specifically, the role of evil in both interpersonal and intergroup hate is explored, and may shed light on the original research question posited here, “if hate protects us from evil, how can it be evil to feel hate?”

### **The Two Hates Problem**

In contemplating the relationship between hate and evil, it seems as if there are two different types of hate or at least two different outcomes of hate. The first protects the self from evil and the second causes evil. However, it is unclear if there is a specific distinction that differentiates the two kinds. In the following sections, the theoretical evidence for different types of hate are considered.

**Interpersonal versus intergroup hate.** Hate serves as a potent example of the difference in intergroup and interpersonal emotions. While intergroup and interpersonal emotions differ from one another (Smith & Mackie, 2015), most publications describing hate invariably lead with a hook about Nazis, Rwanda, or other examples of genocide and hate crimes - horrible atrocities that are no doubt deserving phenomena of our attention and study, but that clearly describe intergroup hate (Allport, 1954; Gaylin, 2003; Staub, 1999; Sternberg, 2003; Sternberg, 2005). Furthermore, most empirical evidence has focused on intergroup hate (Halperin, 2008; Halperin, 2012; Pearson, Dovidio, & Pratto, 2007). It has even been suggested that hate almost exclusively occurs toward groups, not individuals (Gaylin, 2003). But what about the

phenomenon of interpersonal hate, the experience when one person hates another person, not for the group they represent, but for who they are as an individual or what they have done? This phenomenon seems inherently different than intergroup hate. In fact, they seem to demonstrate very different experiences that are covered under the same word. Yet the research on hate very rarely distinguishes them from one another.

**Duplex theory of hate.** According to the duplex theory of hate, hate is comprised of three components (Sternberg, 2003; Sternberg, 2005). The first is the *negation of intimacy*, or repulsion and disgust felt toward the target of hate. The second is *passion*, or the intensity that characterizes hate in the form of anger and fear. The third component is *commitment*, or a cognitive devaluation or dehumanization of the target that manifests as contempt. This theory does not attempt to address the difference in the experiences of intergroup and interpersonal hate, supposing them to be the same, or, more likely, supposing that any differences aren't important to the underlying components.

Another interesting proposal of the duplex theory of hate is that hate incites evil within the self (Sternberg, 2003). The theory argues that hate is the basis of acts of genocide, terrorism, and massacres, and therefore is the basis for human evil. It is further suggested that the roots of evil might collectively form the basis for hate (Baumeister, 1996; Sternberg, 2003). Though the duplex theory of hate proposes that intergroup and interpersonal hate do not differ on this incitation of evil, this idea has not been empirically tested.

**Rational and irrational hate.** Another theory offers a possible dimension by which interpersonal and intergroup hate might be distinguished. Fromm (1973) proffered that there are two kinds of hate: rational and irrational. Hate toward a target in response to personal harm or injustice is considered rational hate. This hate leads to aggression, but with the underlying

function of preserving what is right. In other words, the aggression is justified because it defends the person's life or freedom, as well as promoting the social order by punishing the target.

Hate toward a target because of their race, religion, or group membership, etc., is irrational hate. This irrational or *character-conditioned* hate is considered to be the basis of prejudice, and is based on idealism (Baumeister and Butz, 2005). Irrational hate is not a response to a specific action but is a demonstration of the subject's (rather than the target's) character. This hate is unjustified, poisonous, and involves dehumanization of the target. It does not benefit the self and serves to weaken or destroy social order.

The presence of dehumanization is one of two important points that reflect back to the duplex theory of hate. First, dehumanization is a central tenet of the commitment component of hate (Sternberg, 2005). Irrational hate is based in an ideology in which the target is hated for the characteristics of a group, rather than for his or her own actions (Baumeister & Butz, 2005; Staub, 1999). It is possible that this component is the crux of the difference in interpersonal and intergroup hate and even rational and irrational hate. The ability or willingness to dehumanize others would certainly seem to weaken or destroy social order and would not reflect positively on the subject. Secondly, this type of hate is referred to as "character-conditioned," and refers to the propensity of the person to experience hate toward others based on group membership as an innate part of their character. While the duplex theory of hate suggests that all hate can incite evil in the self and therefore all people have it within them to enact evil if it is activated, this theory suggests that some people have hate within their character and the capacity for evil stems from this.

One interpretation of these theories might suggest that the distinction in rational and irrational hate seems to describe the observed differences in intergroup and interpersonal hate.

Rational hate occurs in response to a personal harm, whereas irrational hate is directed toward groups and group members. Ergo, it suggests that interpersonal hate tends to be rational, whereas intergroup hate tends to be irrational. There are almost certainly exceptions to this statement, but this theory elucidates an important difference. Whatever the distinction, whether intergroup and interpersonal, rational and irrational, or both, there seem to be two different kinds of hate: hate that provides information about the nature of the target of hate and hate that provides information about the character of the person who experiences hate. Importantly, the nature of the information provided by both hates might be the same. The difference is whether the information reveals something about the subject or the target of the hate. The nature of this information is explored in the function that hate serves.

Furthermore, the subjective feeling of hate may not differ. It is possible that they share the components outlined in the duplex theory of hate (Sternberg, 2003). However, it seems that the knowledge about the justification for the emotion underlying each type of hate differs. One hate is internally justified. The subject feels that the hate is necessary in order to protect the self or protect social order. The other hate is not justified. While the subject may feel justified, he or she is not validated in their hate. They are motivated to tear down others for the sake of destruction of the group and of social order.

### **Intergroup vs. Interpersonal Hate: A Morality Issue?**

As hate takes shape as an emotion that is characterized by the revelation of evil in the target and a desire to avoid that target to protect the self and uphold social order, it begins to take on the characteristics of a moral emotion. Moral emotions have been differentiated from nonmoral emotions through several theories (Gray & Wegner, 2011; Haidt, 2003). One theory of moral emotions posits that moral emotions exist along two dimensions: valence, or the presence

of either help or harm, and moral type, or the extent to which the target is an agent or a patient (Gray & Wegner, 2011). According to this theory, moral emotions are elicited by moral events that are observed. These moral events exist within the moral domain, which is determined by a person's moral values (Graham et al., 2011). Additionally, moral emotions have agents, who are responsible for the help or harm, and patients, who receive the help or harm.

A second theory of moral emotions incorporates a self/other distinction in which the person can be an actor in the moral event that elicits the emotion. It is this distinction that most directly applies to interpersonal hate. According to this theory, moral emotions uphold social order through prosocial action tendencies, and are motivated by a desire to benefit others (Haidt, 2003). Though the self may indirectly benefit from the consequences of such emotions, the main recipient is humanity as a whole, or at least someone outside of the self. This theory claims that there are families of moral emotions that either focus on the self or focus on others. For instance, self-conscious emotions, including shame, embarrassment, guilt, and pride, are meant to monitor behavior in the self (Lewis, 1993; Tangney & Fischer, 1995).

The emotion family of most interest to the current research is the other-condemning emotion family, including contempt, anger and disgust. Also called the CAD triad of moral emotions, these emotions monitor the behavior of others (Rozin, Lowery, Imada, & Haidt, 1999). Specifically, the CAD triad (contempt, anger, and disgust) monitor violations of Shweder's "Big Three" moral codes (Shweder, Much, Mahapatra, & Park, 1997). According to the CAD model, anger is elicited in response to violations of personal harm, contempt to violations of an ingroup, and disgust in response to purity or global normative standards. As previously discussed, the duplex theory of hate claims that these very moral violations and emotions are the basis of all hate (Sternberg, 2003; Sternberg, 2005). However, as the duplex theory of hate primarily

concerns intergroup hate, it does not propose that the action tendencies associated with hate are prosocial. If we return to the theory of rational and irrational hate, though, we find that rational hate is concerned with upholding social order. Interpersonal hate, which maps onto rational hate, informs the subject of the evil character and actions of others, and motivates an action tendency of avoidance. In the case of disgust, one of the other-condemning emotions, the action tendency is also avoidance (Haidt, 2003; Rozin et al., 1999). In the case of disgust, avoidance is considered to be prosocial behavior because it ostracizes the target, which removes him or her (and his or her influence) from society. This may also be the case for interpersonal hate. If the target is evil, then he or she is avoided to protect the self and to uphold social order. In other words, interpersonal hate seems to fit snugly within the confines of the other-condemning moral emotions. It is not argued that the same is true for intergroup hate, and the distinction of morality seems to widen the chasm between.

Unlike interpersonal hate, intergroup hate is predicted to precipitate an approach action tendency. In the case of the duplex theory of hate this is even more exaggerated, as the theory is concerned primarily with extreme aggressive acts including genocide, massacre, and terrorism – (Sternberg, 2003). Additionally, the duplex theory of hate asserts that hate is the basis of human evil. This suggests that hate not only monitors the evil of others, but also incites evil within the self. An empirical examination of the duplex theory of hate demonstrates this complexity, though somewhat unintentionally. Examining White Americans' treatment of Black Americans and the role of hate in the U.S. legal system, one study found that hate occurred across the several dimensions but defined evil specifically as dehumanization and an attribution of the group's character (Pearson, Dovidio, & Pratto, 2007). When hate was accompanied by this evil, it significantly predicted people's support of the death penalty. In other words, hate was incited by

a determination about the evil character of the target and in feeling hate the subject condoned violence toward the target. This seems to suggest a sort of “moral exclusion,” such that moral standards that apply to all people are used to determine hate toward a target, but no longer apply to their own behavior toward the target (Opotow, 1990).

To date, the relationship between hate and evil, in which experiencing hate incites evil within the self, has not been explored specifically in interpersonal hate, and there are reasons to think interpersonal hate and intergroup hate differ along this dimension. For instance, while massacre, terrorism, and genocide are explicit acts of violence, an approach action tendency in the extreme, interpersonal hate is characterized by avoidance and self-protection. Such an action tendency does not seem to be the embodiment of evil, as is the argument for intergroup hate. We know from previous research that both interpersonal and intergroup hate involve an attribution that the target is evil. We know also that intergroup hate provokes evil in the self. The question then becomes: does interpersonal hate provoke evil in the self? Evidence says that it should not – avoidance is not evil. However, if people do not differentiate the consequences of interpersonal hate from intergroup hate, this might not be the case. People may assume that feeling hate exposes the self to evil, regardless of the target or action tendency. To fully examine this, it is important to address several points. First, it is important to define evil, and to examine people’s beliefs about evil. More specifically, it is important to determine what people perceive about the relationship between evil and the self, and how the relationship between evil and the self relates to their perceptions and experiences of hate. Finally, it is important to determine whether people are making distinctions between the consequences to the self of interpersonal and intergroup hate. Do people perceive all hate as irrational, even when the hate is justified?

## Statement of the Problem

There are many challenges that arise in the study of hate and evil. One of the most formidable is that both terms are ambiguous and often used without being operationally defined. Research conducted in recent years has done much to supplement our empirical knowledge about the determinants of hate. Hate has been characterized by an attribution of evil about the target (Halperin, 2008; Merrick & Beike, 2019). Though people have indicated that hate informs them about the evil in others, it is not yet understood what criteria people use to determine what is evil. In the current research, three dimensions were identified along which perceptions about evil were explored: intentionality, responsibility, and immorality. Hypothesis 1 is that evil is recognized in immoral acts that are intentional and in which the offending person is responsible for his or her actions. This hypothesis was explored in both Studies 1 and 2, with Study 1 utilizing third-person vignettes portraying immoral acts, and Study 2 utilizing participants' own experiences. In Study 1, intentionality and responsibility were manipulated and participants rated the amount of evil they perceived. In Study 2, participants were asked to relay experiences from their own lives (or the life of someone they knew) and asked to indicate the amount of evil, intentionality, and responsibility they perceived in the offending target. Though participants were asked to provide their beliefs about the nature of evil (i.e. whether it is situational or dispositional or other), this question was exploratory and no specific predictions were offered about this measure. It is important to note that there are no hypotheses related to moral foundation violations. Because each moral foundation violation was only represented by one story, any effect of violation is confounded with story effects. Therefore, though violations are included in subsequent analyses, they will not be discussed in terms of the hypotheses like the other variables will.

A second challenge with the study of hate and evil is understanding the terms of their relationship to one another and whether the same dimensions (responsibility, intentionality, and immorality) are important to both. One way to understand the relationship between hate and evil is to compare it to the relationship between evil and another negative emotion. Recently, we compared the relationship between hate and evil to the relationship between anger and evil. We discovered that hate, but not anger, involves an attribution of evil (Merrick & Beike, 2019). Other researchers have compared the relationship between hate and immorality to that of dislike and immorality, and discovered that hate, but not dislike, involves an attribution of immorality (Van Bavel et al., 2018). Because this second study did not look at the relationship between hate and dislike with evil specifically (using the term *immoral* rather than *evil*), dislike was chosen as the emotion for comparison to hate for the current studies. Van Bavel et al., (2018) demonstrated that hate and dislike differ in degree or intensity and in kind. Therefore, these dimensions were used to compare hate and dislike in the present research.

Hypothesis 2 is that hate and dislike differ in degree. This was tested in both Studies 1 and 2, and in two different ways. In previous research we found that participants did not report intense feelings of hate for third-person scenarios but reported much stronger hate when recounting their own experiences. H2, therefore, is that in Study 1 (in which third-person stories are used), anticipated feelings of dislike will be more intense than anticipated feelings of hate, and in Study 2 (in which personal experiences are used), feelings of hate will be more intense than feelings of dislike. Additionally, in Study 2 hate and dislike were manipulated as independent variables. Since it was hypothesized that hate and dislike differ in degree, then support for H2 was predicted to manifest in the form of main effects of emotion (hate and dislike). Specifically, it was predicted that hate would be considered worse than dislike overall.

Hypothesis 3 is that hate and dislike differ in kind. Support for H3 was predicted to include results that indicated that hate and dislike have differential relationships with evil in both Studies 1 and 2, and by interactions with other variables (subject and target) in Study 2. It was specifically predicted that hate has a stronger relationship with evil than dislike, and that hate differs from dislike, not only in intensity but in the combination of subject, target, and emotional activation (in the form of hate and dislike intensity as outcome variables) it affects.

Another challenge in the study of hate is determining whether people actually use the distinctions that have been outlined here to differentiate different types of hate. While one type of hate is characterized by a tendency to avoid that is motivated by a desire to protect the self, the other type of hate is characterized by a tendency to attack and has been theorized to activate evil within the self. There is evidence, then, that hate not only monitors the evil in others, but incites evil within the subject who feels it. However, it is unclear whether people are making this distinction between different types of hate and what those distinctions are.

To address the hypotheses related to differentiating hate, the data for participants who completed measures about hate in Study 2 were isolated and analyzed without the data for people who answered questions about dislike. It should be noted that these analyses were not originally outlined in the data analysis plan for Study 2, but it was the intention of this Study 2 determine whether different kinds of hate arise when the subject and target of hate is manipulated. Though the analyses were not originally planned, the hypotheses relating to hate were outlined before the data was collected.

Hypothesis 4 is that there are distinct types of hate. There are three sub-hypotheses that outline potential dimensions by which these types of hate might be differentiated that were tested

in Study 2. These dimensions include whether the target of hate is a group or an individual and whether the person experiencing the emotion is the self or someone else.

H4(a) is that hate toward groups will indicate more evil within the subject (self or other) than hate toward individuals, particularly when others experience it (relative to the self). This hypothesis reflects the duplex theory of hate's theoretical assertion that hate incites evil within the self, which occurs most strongly toward groups (Sternberg, 2003). In other words, when others hate groups, they are perceived to be acting on prejudice, and therefore their emotions are less justified.

H4(b) is that others' hate is worse than the self's hate. Specifically, others' hate indicates more activation of evil and is more wrong than hate experienced by the self. In other words, people are more willing to justify their own emotions than the emotions of others. Though this may be a reflection of an attribution bias (e.g., I hate someone because *they* are a bad person, but you hate them because *you* are), it is important in determining the severity of the consequences of hate for both the self and others (Jones & Nisbett, 1987).

H4(c) concerns perceptions about the person toward whom the hate is felt. H4(c) is that the target will be considered more evil when they have offended the self relative to a target who has offended someone else. Once again, this may reflect the self-serving bias, in that what happens to the self is worse than what happens to someone else. Therefore, the offending target of the self's emotions must be worse than the target of someone else's emotions.

Whether or not people make distinctions in types of hate is a particularly important issue. If people do not make this distinction, then by avoiding the harmful consequences of one type of hate people might be missing the protective consequences of the other. To measure the impact of these different dimensions, participants provided feedback about the evil of the target, the evil of

the subject for feeling a negative emotion toward the target, and how ardently the subject resisted feeling the emotion.

The hypotheses are reiterated below:

Hypothesis 1: Hypothesis 1 is that responsibility and intentionality are both necessary for evil to be recognized in another person, but neither is sufficient on its own.

Hypothesis 2: Hate and dislike differ in degree, and feelings of hate are less intense than dislike in response to third-person stories, but feelings of hate and dislike are both more intense in response to hate experiences than dislike experiences.

Hypothesis 3: Hate and dislike differ in kind. Specifically, hate has a stronger relationship with evil and is overall more immoral than dislike.

Hypothesis 4: There are distinct types of hate.

Hypothesis 4a: Hate toward groups indicates more evil within the subject (self or other) than hate toward individuals, particularly when others experience it (relative to the self).

Hypothesis 4b: Others' hate is morally worse than the self's hate.

Hypothesis 4c: The target will be considered more evil when they have offended the self relative to a target who has offended someone else.

## Study 1

### Method

**Participants.** Sample size for Study 1 was predetermined based on a similar previous study to be 400 participants, or 100 participants per cell (Merrick & Beike, 2019). Participants included 273 University of Arkansas students, who completed the online survey through Qualtrics in exchange for partial fulfillment of a research requirement; and 127 Mechanical Turk

workers, recruited through TurkPrime and compensated one dollar and fifty cents for their participation based on a ten-dollar hourly rate. The sample included 197 female, 200 male, one who indicated other, and 12 who chose not to answer. They ranged in age from 18 to 67 with a mean age of 26.14. Eighty nine percent of participants were White, 5% were Black, 4% were Asian, 5% were Latinx, 1% were Native American, 1% were Pacific Islander, and 1% indicated other. Participants could choose more than one ethnicity. Per the data analysis plan, any participants who missed more than four comprehension check questions would be excluded. However, after data collection it was discovered that more than two-thirds of the sample should be excluded based on this criterion. Furthermore, closer inspection revealed that all but ten participants (out of 201) in the Not Responsible conditions were excluded under the criterion, and there were no other differences in gender, race, or age between the participants that were to be included and those that were to be excluded. It was therefore obvious that participants in the Not Responsible condition had either misunderstood the information in the vignettes, that the vignettes themselves were not successfully manipulated in this condition, or that participants did not understand the questions. Further analyses (presented in the Results and Discussion section) revealed a failure to manipulate responsibility in the vignettes, or at least failure of participants to understand what was asked of them in the comprehension and manipulation check questions. Because such a large portion of the sample was excluded, the data analyses were conducted in two ways: with only the participants who met the original inclusion criteria, and with all participants included. The results of the analyses with all of the participants were presented in the text. The results of the diminished sample are available in Appendix A<sup>1</sup>.

---

<sup>1</sup> The only differences in the results of Study 1 with and without the excluded participants were effects of responsibility that occurred in the larger sample. Those effects were expectedly absent from the smaller dataset, as almost all of the participants in the Not Responsible condition were not included.

**Procedure.** Study 1 was conducted online. Participants first completed an informed consent which was recorded on a separate survey from the rest of their responses to preserve anonymity. Participants read six short vignettes and respond to questions corresponding to each. Study 1 employed a 2 (responsible, not responsible)  $\times$  2 (intentional, unintentional)  $\times$  6 (moral foundations: harm, fairness, ingroup, authority, purity, liberty) mixed model design, with responsibility and intentionality presented between subjects and the moral foundations presented within subjects. After completing the informed consent, participants were assigned to one of four conditions in which the actions of the character in the vignette varied in intentionality and responsibility. These groups included responsible and intentional, responsible but unintentional, not responsible but intentional, and not responsible and unintentional (see Appendix B for the materials for Study 1). Within each condition, participants read vignettes portraying each of the moral foundations. All of the characters were male to prevent any gender differences in participants' perceptions. Presentation order of the vignettes was randomized for each participant to prevent order effects.

Each vignette included one named character (hereafter referred to as the Focal Character) and one unnamed character that was an acquaintance of the Focal Character (hereafter referred to as the Associate). In each vignette, the Associate encouraged the Focal Character to commit a moral violation. The Focal Character was the main character of the story. The actions of the Focal Character were manipulated in intentionality and responsibility, constituting the four conditions described above. Each vignette ended with statements about the responsibility and intentionality of the actions of the Focal Character to minimize participant confusion and to ensure the manipulation was obvious.

After each vignette, participants answered two true/false questions that served as comprehension checks. These included two statements about the intentionality and responsibility of the Focal Character and reflected the two final statements of the vignette which the participants had to indicate as either true or false. Participants were excluded if they missed more than four of these questions across all of the comprehension questions. For each vignette, participants then completed a short survey in which they indicated their agreement with several statements using a seven-point scale from 1 (*very strongly disagree*) to 7 (*very strongly agree*). The first set of statements concerned participants' perceptions about the Focal Character beginning with intentionality and responsibility (e.g. "Tom acted on purpose;" "Tom is responsible for his actions.>"). These statements were similar to the attention check questions, but rather than checking participants' comprehension of the vignettes, these questions were intended to measure participants' beliefs about the intentionality and responsibility of the Focal Character's actions and were prefaced by asking participants to what extent they *believed* these statements to be accurate. Next, participants indicated their agreement with statements that the Associate and the Focal Character were evil (e.g. "Tom is evil;" "Tom's friend is evil.>"). once again using a seven-point scale from 1 (*very strongly disagree*) to 7 (*very strongly agree*). The item about the evil of the Associate was intended to mask the focus of the study and was not analyzed.

Following these statements, participants answered questions regarding their own hypothetical emotions had they been the victims/observers in the scenario described in the vignette on a scale from 1 (*none at all*) to 7 (*a great deal*). Once again, there were questions about both characters but only the questions regarding the Focal Character were of interest. Specifically, participants were asked to indicate the amount of hate and dislike they would feel

for both characters had they been victims of their actions (in vignettes that included victims) or observers of their actions (in vignettes that did not include victims). For instance, in the case of harm violations, the question was “If you were the victim of this situation, how much hate/dislike would you feel toward [Focal Character]?” In the case of purity violations (which were less likely to have victims) an example item was “If you saw this happen, how much hate/dislike would you feel toward [Focal Character]?” These items were designed to capture the relationship between perceived evil/immorality and the emotions of hate and dislike by measuring participants’ anticipated hate or dislike if they were in the situation. To clarify this, in the Results section the hate and dislike analyses are referred to as “anticipated hate/dislike.”

Finally, for each vignette participants were asked to provide their own perspective of the Focal Character’s morality in a series of agreement statements on a scale from 1 (*very strongly disagree*) to 7 (*very strongly agree*), including items such as “[Focal Character] is infected by [Associate]’s immorality” and “[Focal Character] chose to act immorally.” These items served to indicate participants’ beliefs about the contagion or activation of evil in the Focal Character after he had been exposed to the moral violations (immorality) of the Associate. These questions were not included in the analyses of the present research.

After participants responded to all six vignettes, they completed several personality and beliefs measures. These measures were not included in the analyses of this study but will be analyzed at a later time. These measures included the Moral Vitalism Scale (Bastian et al., 2015), the Moral Foundations Questionnaire (Graham & Haidt, 2011), items regarding beliefs about evil, and items about consequences for bad people. Finally, participants completed demographic measures and were debriefed.

## Data Analysis Plan

Any participants who missed more than four of the comprehension check questions were planned to be excluded from the analysis. All remaining participants were to be included.

Hypotheses 1, 2, and 3 were addressed in Study 1. The analysis plan is discussed in terms of the hypotheses.

Hypothesis 1 is that responsibility and intentionality are both necessary for evil to be recognized, but neither is sufficient on its own. A 2 (responsible, not responsible)  $\times$  2 (intentional, unintentional)  $\times$  6 (moral foundations: harm, fairness, ingroup, authority, purity, liberty) mixed model analysis of variance was conducted on the item measuring the perceived evil of the Focal Character (whose actions were manipulated). Based on H1, it was predicted that the analysis would reveal an interaction between responsibility and intentionality of the Focal Character. Specifically, it was predicted that a post-hoc simple effects test would reveal that Focal Characters in vignettes that displayed both intentionality and responsibility would be considered significantly more evil than vignettes with any other combination of responsibility and intentionality. Because H1 predicted that both intentionality and responsibility were necessary for the recognition of evil, the main effects of intentionality and responsibility were not considered important. However, it was considered likely that two main effects would emerge from this analysis. First was a main effect of responsibility is likely, such that the Focal Character's actions would be perceived as more evil when the character was directly responsible for his actions. The second was a main effect of intentionality, such that the Focal Character's actions would be perceived as more evil when they were intentional. No main effect was predicted for moral foundation, based on MFT's suggestion that evil exists at all levels of moral

foundations (Graham & Haidt, 2014). However, as mentioned previously any effects of violation might also be attributed to story effects and therefore were not reported in the main text.

An additional analysis to explore H1 consisted of computing participants' average scores of intentionality of the Focal Characters' actions, responsibility for the Focal Characters' actions, and perceived evil of the Focal Characters across all vignettes, and then conducting a simultaneous regression in which participants' intentionality and responsibility ratings of each vignette were entered as predictors of evil. It was hypothesized that intentionality and responsibility would independently predict participants' perceptions of evil of the Focal Character.

Hypothesis 2 is that hate and dislike differ in degree. H2 and partially H3, that hate and dislike differ in kind, were explored through the same analysis. A 2 (responsible, not responsible)  $\times$  2 (intentional, unintentional)  $\times$  6 (moral foundations: harm, fairness, ingroup, authority, purity, liberty)  $\times$  2 (emotion type: anticipated hate, anticipated dislike) mixed model analysis of variance was conducted on anticipated emotion. Support for H2 would include a main effect of emotion, such that anticipated feelings of dislike would be more intense than anticipated feelings of hate. Regarding H3, a two-way interaction of emotion and responsibility was expected, in which hate, but not dislike increased when the Focal Character was responsible, compared to when he was not responsible. A similar two-way was predicted between emotion and intentionality, in which hate, but not dislike, increased when the Focal Character's actions were intentional, compared to when they were unintentional.

Support for H3 was also explored through a simultaneous regression. Average anticipated hate and average anticipated dislike scores were computed and entered as simultaneous predictors of participants' average perceptions of evil of the Focal Character. It was predicted

that hate, but not dislike, would be a strong predictor of perceptions of evil of the Focal Character.

## **Results and Discussion**

The first analyses presented for Study 1 are manipulation checks of the intentionality and responsibility portrayed in each of the conditions. These analyses were not presented as part of the data analysis plan. However, in preparing the data for analysis the comprehension check questions revealed that participants might have misinterpreted the comprehension questions (particularly the responsibility question), and therefore had misunderstood the manipulation check question (which were the same as the comprehension questions, except that they incorporated participants' beliefs about the responsibility and intentionality of the Focal Character). Additionally, these analyses were performed to provide a potential interpretation for some unexpected results from the planned analyses. They were presented first because they inform the analyses that follow and therefore it is important to understand them before moving forward.

The first of these manipulation checks was performed on intentionality. After each scenario, participants were asked to indicate their belief that the Focal Character acted intentionally (this question was separate from the true/false comprehension check question which asked participants whether the act was intentional based on the information in the question). A composite score was calculated, averaging participants' intentionality belief scores across the moral violations. A t-test comparing the intentional and unintentional conditions was conducted on the composite of participants' intentionality belief ratings, revealing that intentional vignettes ( $M = 6.27$ ;  $SE = .06$ ) were perceived as more intentional than unintentional vignettes ( $M = 4.12$ ,  $SE = .07$ ), indicating that the manipulation of intentionality was successful,  $t(396) = 24.06$ ,  $p <$

.001. A similar composite score was calculated for participants' beliefs that the Focal Character was responsible for their actions. A t-test was conducted on this composite score comparing responsible and not responsible conditions, revealing that responsible vignettes ( $M = 5.59$ ,  $SE = .08$ ) were not perceived as significantly different than vignettes in which the character was not responsible ( $M = 5.46$ ,  $SE = .09$ ), indicating that the manipulation of responsibility was not successful, or that the question assessing this dimension did not adequately represent the intended dimension. In summary, of the two between subject variables, only intentionality was successfully manipulated and assessed. However, as will be demonstrated in the following analysis, responsibility was not rendered obsolete.

Before exploring the analyses for Study 1, it is noteworthy to mention that, while there were no predictions for effects of violation type for any of the analyses, multiple effects were observed throughout Study 1. These effects might signify a hierarchy for the differing violations, however, there are two important reasons to be cautious in accepting such an interpretation. The first is that each violation was represented by only one vignette. Therefore, the differences could be attributed to the stories themselves rather than the type of violation portrayed. The second is evidence from our own previous research. Across four different studies, we found that the violation type associated with hate was variable – in two studies it was based on harm and liberty violations, in another it was based on group and fairness violations, and in another it was based on purity violations (Merrick & Beike, 2019). In that research, the type of violation did not seem to matter as much as the perception that the offender was evil and the victim was motivated to avoid them. Therefore, while the effect of violation type might signify something meaningful, it could also be subject to change. In the interest of transparency, the results of these analyses,

including the effects of violation type are available in Tables 1-4 but they will not be otherwise reported here.

Hypothesis 1 is that responsibility and intentionality are both necessary for evil to be recognized, but neither is sufficient on its own. A 2 (responsible, not responsible)  $\times$  2 (intentional, unintentional)  $\times$  6 (moral foundations: harm, fairness, ingroup, authority, purity, liberty) mixed model analysis of variance was conducted on the item measuring the perceived evil of the Focal Character (whose actions were manipulated). Contrary to predictions, there was no significant interaction between intentionality and responsibility of the Focal Characters actions. There was a main effect of intention, such that the Focal Characters in vignettes portraying intentional acts ( $M = 4.01$ ,  $SE = .07$ ) were perceived to be more evil than in vignettes portraying unintentional acts ( $M = 2.38$ ,  $SE = .07$ ),  $F(1, 394) = 262.24$ ,  $p < .001$ ,  $\eta_p^2 = .40$ . There was also a main effect of responsibility, such that the Focal Character in vignettes portraying the character as responsible for his actions ( $M = 3.37$ ,  $SE = .07$ ) were considered more evil than in vignettes portraying the character as not responsible for his actions ( $M = 3.02$ ,  $SE = .07$ ),  $F(1, 394) = 12.05$ ,  $p = .001$ ,  $\eta_p^2 = .03$ . The prediction that both intentionality and responsibility are necessary for the perception of evil was not supported. However, the main effects of both intentionality and responsibility signifies that they may each be sufficient for the recognition of evil. Interestingly, despite the previous analysis revealing that the responsibility manipulation was not effective, in perceiving the evil of the Focal Character, participants did discern a difference in the vignettes in which the character was portrayed as responsible and those in which he was not. These results tend to support the argument that participants did not understand the question, rather than a failure of the concept.

An additional analysis to explore H1 consisted of computing participants' average scores of intentionality of the Focal Characters' actions, responsibility for the Focal Characters' actions, and perceived evil of the Focal Characters across all vignettes. A simultaneous linear regression was then conducted that demonstrated that intentionality and responsibility were both significant predictors of evil,  $F(2, 395) = 102.22, p < .001$  with an  $R^2 = .34$ . Intentionality was a significant predictor of evil ( $\beta = .37, p < .001$ ), as was responsibility ( $\beta = .22, p = .03$ ). Again, because responsibility failed as a manipulation, it was unclear what (if any) differences participants perceived in those scenarios relative to the not responsible ones, but it is clear that the more participants believed the Focal Character was responsible for an immoral act, the more evil they perceived him to be.

H2 is that hate and dislike differ in degree. A 2 (responsible, not responsible)  $\times$  2 (intentional, unintentional)  $\times$  6 (moral foundations: harm, fairness, ingroup, authority, purity, liberty)  $\times$  2 (emotion type: anticipated hate, anticipated dislike) mixed model analysis of variance was conducted on anticipated emotion. Participants' overall anticipated feelings of hate and dislike toward the Focal Character were different from each other. As predicted for Study 1, anticipated dislike ( $M = 4.51, SE = .06$ ) was more intense overall than anticipated hate ( $M = 3.55, SE = .07$ ),  $F(1, 382) = 290.46, p < .001, \eta_p^2 = .43$ . To understand these results, context is potentially crucial. The questions regarding hate and dislike asked participants to report how much hate and dislike they would experience as the victim of the story. As will be demonstrated in Study 2, people report resisting hate more than dislike and feeling that hate is more immoral than dislike, so it is possible that people thought they would experience less hate in a third-person scenario because they perceived feeling hate to be wrong. Given potential concerns about

the social desirability of the question and people's ability to control their emotions in imagined stories, the fact that people reported feeling more dislike than hate makes sense.

H3 is that hate and dislike differ in kind. H3 was explored through the previous analysis that also provided evidence for H2, as well as a simultaneous regression. First, the 2 (responsible, not responsible)  $\times$  2 (intentional, unintentional)  $\times$  6 (moral foundations: harm, fairness, ingroup, authority, purity, liberty) mixed model analysis of variance conducted on the item measuring anticipated feelings of hate and anticipated feelings of dislike was examined. Contrary to predictions, there was no two-way interaction between emotion and responsibility of the target (which is not surprising, given the initial analysis that showed no effect of the responsibility manipulation question), nor was there a two-way interaction between emotion and intentionality of the target.

Support for H3 was also sought through a simultaneous linear regression which revealed that both anticipated feelings of hate and anticipated feelings of dislike were significant predictors of perceptions of evil of the target,  $F(2, 317) = 103.63, p < .001$  with an  $R^2 = .40$ . As predicted, anticipated hate was a significant predictor of evil ( $\beta = .24, p < .001$ ). Contrary to predictions, however anticipated dislike also predicted evil ( $\beta = .39, p < .001$ ). Though this does not align fully with the predictions of H3, it is important to note that both anticipated hate and anticipated dislike independently predicted perceptions that the target was evil, indicating that hate and dislike differentially contribute to people's notions about the presence of evil.

To review, Study 1 provided partial support for H1, by revealing that people use intentionality and responsibility as sufficient dimensions by which they perceive evil, though they were not mutually necessary, as predicted. This support is complicated, however, by the fact that the manipulation or assessment of responsibility failed. H2 was supported by demonstrating

that people anticipated stronger feelings of dislike than hate in third-person scenarios. H3 was somewhat supported. Anticipated feelings of hate and dislike did not differ as expected in terms of intentionality and responsibility, as predicted, and hate did not have a stronger or singular relationship with evil. However, hate and dislike were revealed to have independent relationships with evil. Therefore, while it is unclear how they differ in kind, it is clear that they do.

There were two goals of Study 2. First, Study 2 attempted to replicate the results Hypotheses 1, 2, and 3 from Study 1 but with more external validity: participants provided their own experiences of hate and dislike (or the experiences of someone they knew) to determine if people use the same dimensions to decide who is evil. Second, Study 2 sought to answer the theoretically contradictory question about the relationship between hate and evil more closely – if hate informs us about evil, then how can it be evil to feel hate? This question was examined by manipulating different dimensions by which people might differentiate the positive and negative outcomes of hate, including the subject (self, other), the target (group, individual), and comparing these dimensions for hate and dislike to determine if and how the distinctions of hate differed from those of dislike. These dimensions were then examined as they related only to hate, to distinguish whether different kinds of hate were observed.

## Study 2

### Method

**Participants.** Eight hundred and twenty-five participants were recruited for the study. Of those, 16 failed to begin the experiment and nine were excluded based on the a priori exclusion criteria, by answering the question about their hate or dislike experience with nonsense (e.g., “2;” “FFF”). The final sample for Study 2 included 141 University of Arkansas students who completed the online survey through Qualtrics in exchange for partial fulfillment of a research

requirement; and 660 Mechanical Turk workers recruited through TurkPrime and compensated one dollar for their participation based on a ten-dollar hourly rate. The participants included 383 females, 395 males, three who indicated other, and 20 who chose not to reveal their gender. They ranged in age from 18 to 75 with a mean age of 34.60. Eighty percent of participants were White, 7% were Black, 8% were Asian, 6% were Latinx, 1% were Native American, 1% were Pacific Islander, and 1% indicated other. Participants could choose more than one ethnicity. Sample size was predetermined at 800 participants, or 100 participants per cell.

**Procedure.** Study 2 was completed online. Participants first completed an informed consent which was recorded on a separate survey from the rest of their responses to preserve anonymity. After beginning the survey, participants were prompted to provide an example of an emotional experience and respond to corresponding questions. This study employed a 2 (emotion: hate, dislike)  $\times$  2 (subject: self, other)  $\times$  2 (target: group, individual) between subjects design. Participants described an example of a time they or someone they knew experienced either hate or dislike toward a group or individual. In the Other condition participants also provided the name of the person who experienced either hate or dislike. The Other's name was then piped into the following questions to minimize participant confusion.

After providing and describing an experience, participants answered a series of questions. The first set of questions included questions about the subject's experience and perceptions about the target (either a group or a person). In the Self conditions, these questions asked for the intensity of participants' own emotions and judgments (e.g. "How intensely did you feel hate toward the person?"). In the Other conditions, the questions asked for the intensity of the Other's emotions and judgments (e.g. "How intensely did you perceive the person to feel hate toward the person?"). These questions were answered on a scale from 1 (*not at all*) to 7 (*a great deal*). An

example is included below that demonstrates the items for the Self/Group/Hate condition. The questions pertaining to threat and immorality of the target were not analyzed for the current study but will be explored at a later date. Participants indicated the intensity of the following items, either about the self or the person they witnessed experiencing the indicated emotion:

“hate toward the group”

“dislike toward the group”

“threatened by the group”

“the group was evil”

“the group did something on purpose”

“the group had a choice about their actions”

“the group was immoral”

The next set of questions pertained to the morality of the subject of the emotional experience (either self or other) and measured the possible consequences of experiencing hate or dislike toward a group or person. These items were matched to emotion condition, so that the items in the dislike condition asked about dislike (e.g. “I resisted feeling dislike toward the group”) and items in the hate condition asked about hate (e.g. “I resisted feeling hate toward the group”). An example of the items is included below that demonstrates the items for the Self/Group/Hate condition. The first three items of this set of questions were filler questions and were not subject to analysis. Participants rated their agreement with to the following items on a scale from 1 (*very strongly disagree*) to 7 (*very strongly agree*):

“By experiencing hate toward the group, I activated evil in myself”

“By experiencing hate toward the group, I could have infected others with evil”

By experiencing hate toward the group, I could have harmed others”

“I resisted feeling hate toward the group”

“It was wrong for me to feel hate toward the group”

After answering questions about the target, about the subjective experience, and about the consequences of experiencing hate or dislike on the subject, participants indicated their desired action tendency. This was done through two different questions. The first question that addressed action tendency included two slider scales from one to 100 that asked participants to indicate their present desire to approach the target and avoid the target. The second set of items asked participants how much they would like to engage in the action tendency (approach or avoid) at the time of the experience, and asked them to specify the motivation for the action tendency, of which there were three options: to harm the target, to protect the self from further harm, or to avoid consequences. An example of this question is available below for the Self/Group/Hate condition. Participants indicated their desire to execute each of the following actions from 1 (*not at all*) to 7 (*a great deal*):

“confront or attack this group because you wanted to cause them pain or harm”

“confront or attack this group because you wanted to protect yourself from further harm”

“confront or attack this group because you wanted to avoid the consequences of staying away from them”

“stay away from this group because you wanted your absence to cause them pain or harm”

“stay away from this group because you wanted to protect yourself from further harm”

“stay away from this group because you wanted to avoid the consequences of confronting or attacking them”

After participants described the experience and answered the corresponding questions, they answered the same personality and beliefs measures from Study 1. Once again, these measures were not included in the analyses of this study but will be analyzed at a later time. These measures included the Moral Vitalism Scale (Bastian et al., 2015), the Moral Foundations Questionnaire (Graham et al., 2011), items regarding beliefs about evil, and items about consequences for bad people. Finally, participants completed demographic measures and were then debriefed and paid (Mechanical Turk workers) or administered research credit (students) for participation. The materials for Study 2 are available in Appendix C.

### **Data Analysis Plan**

The measures of interest for Study 2 included the intensity of hate, the intensity of dislike, desire to approach the target, desire to avoid the target, the evil of the target, the wrongness of experiencing the emotion, the resistance to feeling the emotion, the intentionality of the target, and the responsibility of the target. Evil of the target, wrongness of the emotion, and resistance to the emotion were each subject to a 2 (emotion experience: hate, dislike)  $\times$  2 (subject: self, other)  $\times$  2 (target: group, individual) between subject analysis of variance. Emotion intensity (intensity of hate and dislike) was treated as a within subjects variable, and entered into the same 2 (emotion experience: hate, dislike)  $\times$  2 (subject: self, other)  $\times$  2 (target: group, individual)  $\times$  2 (emotion intensity: hate, dislike) mixed model analysis of variance. Desire to approach and avoid the target were also treated as a within subjects variable, and entered into a 2 (emotion experience: hate, dislike)  $\times$  2 (subject: self, other)  $\times$  2 (target: group, individual)  $\times$  2 (action tendency: approach, avoid) mixed model analysis of variance. These analyses are described in terms of the hypotheses.

Hypothesis 1, that evil is recognized in immoral acts that are intentional and in which the offending person is responsible for his or her actions, was addressed with a regression analysis similar to one from Study 1. Participants' ratings of the intentionality and responsibility of the target were entered as simultaneous predictors of their perception that the target was evil. It was hypothesized that both intentionality and responsibility would predict participants' perceptions of evil of the target.

Hypothesis 2 is that hate and dislike differ in degree. Support for H2 would include main effects of emotion experience for any of the outcome measures of interest. Specifically, the target of hate was predicted to be more evil, more responsible for their actions, more intentional than the targets of dislike, and hate was predicted to have a stronger action tendency than dislike. Additionally, hate experiences were predicted to be wrong to experience, more resisted, and to activate higher intensity of negative emotions (hate and dislike) than dislike experiences.

Hypothesis 3 is that hate and dislike differ in kind. There were two types of evidence intended to support H3. The first was a simultaneous regression in which it was predicted that hate, but not dislike, was a strong predictor of perceptions of evil of the target, though based on the results of Study 1, it would not be surprising if dislike was also a predictor of perceptions of evil of the target. If it was found that hate and dislike are both predictors of evil, it would support H2, that hate and dislike differ in intensity.

Support for H3 would also include any two-way interactions in the analyses of variance between emotion experience (hate, dislike) and subject (self, other), emotion experience and target (individual, group), or emotion experience and reported emotional intensity (hate, dislike) would indicate a difference in kind between hate and dislike for the following outcome measures: evil of the target, wrongness of feeling the emotion, resistance to feeling the emotion,

hate and dislike intensity, and action tendency. The specific nature of these are covered under subsequent predictions.

First, the  $2$  (emotion: hate, dislike)  $\times$   $2$  (subject: self, other)  $\times$   $2$  (target: group, individual) analysis of variance conducted on participants ratings of evil of the target was anticipated to produce a two-way interaction between emotion and subject and target. Mean comparison was predicted to reveal that the target of hate experienced by the self was considered more evil than dislike experienced by the self or the target of any of the other's experiences.

Next, the  $2$  (emotion experience: hate, dislike)  $\times$   $2$  (subject: self, other)  $\times$   $2$  (target: group, individual) analysis of variance conducted on participants ratings of how wrong it was to experience the emotion was expected to produce a three-way interaction between emotion experience, subject, and target. Mean comparison was expected to reveal that others' hate toward groups was more wrong than their dislike toward any target, across all levels of subject. A two-way interaction was predicted between emotion and subject, such that others' experiences of hate were more wrong than their experiences of dislike or any of the self's experiences. A two-way interaction was also predicted between emotion and target, such that hate toward groups is more wrong than dislike toward either target.

Next, the  $2$  (hate, dislike)  $\times$   $2$  (self, other)  $\times$   $2$  (group, individual) analysis of variance conducted on participants' ratings of how strongly the subject resisted the negative emotion was expected to produce a two-way interaction between emotion and subject. Mean comparison was predicted to reveal that the self is more likely to resist hate than dislike, and any emotion that the other experiences.

Next, the  $2$  (emotion experience: hate, dislike)  $\times$   $2$  (subject: self, other)  $\times$   $2$  (target: group, individual)  $\times$   $2$  (emotion intensity: hate intensity, dislike intensity) mixed model analysis

of variance was predicted to produce a two-way interaction, such that both hate and dislike intensity would be higher in hate experiences than dislike experiences.

Finally, the  $2$  (emotion experience: hate, dislike)  $\times$   $2$  (subject: self, other)  $\times$   $2$  (target: group, individual)  $\times$   $2$  (action tendency: approach, avoid) mixed model analysis of variance was predicted to produce a two-way interaction, such that hate would have a stronger action tendency (and specifically, a stronger avoidance tendency) than dislike. It was subsequently predicted that hate would have a stronger action tendency than dislike and that both emotions would be characterized by a desire to avoid, rather than approach.

## **Results and Discussion**

**Planned Analyses.** The measures of interest for Study 2 included the intensity of the emotions, the evil of the target of the experience, the wrongness of experiencing the emotion, the resistance to feeling the emotion, the intentionality of the target, and the responsibility of the target. Each of the dimensions above was subject to a  $2$  (hate, dislike)  $\times$   $2$  (self, other)  $\times$   $2$  (group, individual) between subject analysis of variance. It is important to point out that hate and dislike intensity were measured as dependent variables as well as being independent conditions of the study. Participants described instances of either hate or dislike but were asked to provide the intensity of both emotions, regardless of condition. Additional analyses included two regressions similar to those conducted in Study 1: one in which hate and dislike intensity were used to predict evil of the target, and another in which intentionality and responsibility were used to predict evil of the target. The results of the analyses were discussed in terms of the specific hypotheses rather than in order of the analyses for clarity. Because the analyses were discussed in terms of the hypotheses, the results of interest for each of the analyses of variance are the main effects of emotion type and any interactions between emotion type and the other variables (as

outlined in the data analyses plan). Therefore, while there are many main effects of the other variables throughout the analyses, these are not discussed in the results in order to reduce cognitive load on the reader. Analysis of variance tables, including the main effects that are not discussed in the results, are available in Tables 5-9 so that the analyses can be viewed in complete form.

Hypothesis 1, that evil is recognized in immoral acts that are intentional and in which the offending person is responsible for his or her actions, was addressed with a simultaneous regression analysis similar to one from Study 1. Participants' ratings of the intentionality and responsibility of the target were entered as simultaneous predictors of their perception that the target was evil, which was found to be significant,  $F(2, 783) = 107.55, p < .001$  with an  $R^2 = .22$ . As predicted, intentionality was a strong predictor of evil ( $\beta = .50, p < .001$ ), while responsibility was only a marginally significant predictor of evil ( $\beta = .10, p = .07$ ). This analysis replicates the results from Study 1 and once again suggests that intentionality seems very important to the perception of evil, while responsibility is less so. Once again, it is unclear whether it was the language that confused participants (i.e., does everyone understand the word responsibility the same way?) or if there is another dimension that is important to the attribution of evil. In Study 1, perceptions of evil were greater in intentional and responsible vignettes, though the responsibility manipulation failed. This, along with the regression analyses suggest that responsibility is either misunderstood or is not related to evil. The fact that the Focal Characters in the responsible vignettes were still considered more evil than the ones that were not suggests that there is a dimension that is important to participants, but that they do not recognize it as responsibility.

Hypothesis 2 is that hate and dislike differ in degree. Support for H2 would include main effects of emotion experience for any of the outcome measures of interest. A 2 (hate experience, dislike experience)  $\times$  2 (self, other)  $\times$  2 (group, individual) analysis of variance was conducted on each of the outcome measures. Results included main effects of emotion for four outcome variables that revealed differences in the intensity of hate and dislike. Participants reported that the target of emotional experiences were more evil for hate ( $M = 4.36$ ,  $SE = .10$ ) than dislike ( $M = 3.65$ ,  $SE = .10$ ),  $F(1,778) = 24.64$ ,  $p < .001$ ,  $n_p^2 = .03$ . They also reported that it was more wrong to experience hate ( $M = 3.61$ ,  $SE = .10$ ) than dislike ( $M = 2.99$ ,  $SE = .10$ ),  $F(1,778) = 21.33$ ,  $p < .001$ ,  $n_p^2 = .03$ , and relatedly that they were more likely to resist hate ( $M = 2.93$ ,  $SE = .08$ ) than dislike ( $M = 2.49$ ,  $SE = .08$ ),  $F(1,778) = 15.22$ ,  $p < .001$ ,  $n_p^2 = .02$ . Another analysis of interest included a 2 (hate experience, dislike experience)  $\times$  2 (self, other)  $\times$  2 (group, individual)  $\times$  2 (emotion type: anticipated hate, anticipated dislike) mixed model analysis of variance was conducted on anticipated emotion. For this analysis, hate and dislike experiences were once again manipulated as an independent variable, but they were also measured as an outcome variable. Therefore, two main effects were predicted for this analysis. One of the emotion experience (hate and dislike) and the other for the intensity of the emotions that accompanied that experience (hate and dislike). There was a main effect of condition, such that hate experiences ( $M = 5.71$ ,  $SE = .07$ ) were accompanied by more intense negative emotions than dislike experiences ( $M = 5.10$ ,  $SE = .07$ ),  $F(1,778) = 40.02$ ,  $p < .001$ ,  $n_p^2 = .05$ . There was also a main effect of emotion intensity, such that dislike intensity ( $M = 5.95$ ,  $SE = .05$ ) across all emotional experiences was more intense than hate intensity ( $M = 4.86$ ,  $SE = .06$ ),  $F(1,778) = 474.07$ ,  $p < .001$ ,  $n_p^2 = .38$ . Together, these main effects suggest that hate experiences are characterized by more intense negative emotions, but that the subjective feeling of dislike is

more intense than the feeling of hate. There were no main effects of emotion for the other variables, including intentionality of the target and responsibility of the target.

Additional evidence that hate and dislike differ in degree was observed in the analyses regarding action tendency. A 2 (hate experience, dislike experience)  $\times$  2 (self, other)  $\times$  2 (group, individual)  $\times$  2 (approach, avoid) mixed model analysis of variance was conducted. First, there was a main effect of action tendency, such that both hate and anger were characterized by avoidance ( $M = 71.82, SE = 1.12$ ), rather than approach ( $M = 28.18, SE = 1.11$ )  $F(1, 778) = 578.82, p < .001, \eta_p^2 = .43$ . More interestingly though, there was a main effect of emotion such that hate experiences ( $M = 51.63, SE = .92$ ) had a stronger action tendency than dislike experiences ( $M = 48.37, SE = .91$ ),  $F(1, 778) = 6.27, p = .01, \eta_p^2 = .008$ . Though both means were close to the midpoint, hate had a significantly stronger action tendency than dislike which differentiates hate and dislike in degree.

Hypothesis 3 is that hate and dislike differ in kind. There were two types of evidence intended to support H3. The first was a simultaneous regression in which it was revealed that hate and dislike were both predictors of evil  $F(2, 783) = 165.00, p < .001$  with an  $R^2 = .30$ . Hate was a significant predictor of evil ( $\beta = .48, p < .001$ ), as predicted. Contrary to predictions (but consistent with Study 1), dislike was also a significant predictor of evil ( $\beta = .29, p < .001$ ). Though dislike was not anticipated to be a predictor of evil, it is important to note that both hate and dislike independently predicted perceptions that the target was evil, indicating that hate and dislike differentially contribute to people's notions about the presence of evil.

As argued in the data analysis plan, differences in hate and dislike in kind would be demonstrated by two-way interactions in the analyses of variance that were proposed. There were no differences in kind between hate and dislike for perceptions that the target was evil,

intentionality of the target, responsibility of the target, or that the subject resisted feeling the emotion. There were interactions of interest in the 2 (hate experience, dislike experience)  $\times$  2 (self, other)  $\times$  2 (group, individual)  $\times$  2 (hate intensity, dislike intensity) mixed model analysis of variance. First, there was a two-way interaction between hate and dislike intensity (the within subjects measures of hate and dislike) and subject (self, other),  $F(1, 778) = 8.69, p < .001, \eta_p^2 = .01$ . An analysis of simple effects with a Bonferroni correction revealed that the self experienced more intense dislike ( $M = 6.02, SE = .06$ ) than hate ( $M = 4.77, SE = .08$ ),  $F(1, 778) = 325.03, p < .001, \eta_p^2 = .30$ . Others also experienced more intense dislike ( $M = 5.89, SE = .07$ ) than hate intensity ( $M = 4.94, SE = .09$ ),  $F(1, 778) = 167.18, p < .001, \eta_p^2 = .18$ . There was not a significant difference in the experience of the self and other. There was also a two-way interaction between emotion intensity (the within subjects measures of hate and dislike) and emotion experience (the between subjects measures of hate and dislike),  $F(1, 778) = 52.81, p < .001, \eta_p^2 = .06$ . An analysis of simple effects with a Bonferroni correction revealed that hate intensity was significantly higher in hate experiences ( $M = 5.34, SE = .09$ ) than in dislike experiences ( $M = 4.37, SE = .09$ ) as expected,  $F(1, 778) = 64.02, p < .001, \eta_p^2 = .08$ . Dislike intensity was also higher in hate experiences ( $M = 6.07, SE = .07$ ) and dislike experiences ( $M = 5.84, SE = .07$ ),  $F(1, 778) = 6.53, p = .01, \eta_p^2 = .008$ .

Additional evidence that hate and dislike differ in kind was observed in the analyses regarding action tendency. A 2 (hate experience, dislike experience)  $\times$  2 (self, other)  $\times$  2 (group, individual)  $\times$  2 (approach, avoid) mixed model analysis of variance revealed a two-way interaction of action tendency and emotion,  $F(1, 778) = 7.10, p = .008, \eta_p^2 = .009$ . An analysis of simple effects with a Bonferroni correction revealed that hate was significantly more likely to have an approach action tendency ( $M = 32.33, SE = 1.58$ ) than dislike ( $M = 24.14, SE = 1.56$ ),

$F(1, 778) = 13.28, p < .001, \eta_p^2 = .02$ . There was no significant difference in hate and dislike for avoidance action tendency.

Collectively, the analyses comparing hate and dislike revealed some differences between the two, which are not just based on intensity, but also on kind, as the results of Van Bavel et al. (2018) suggested, though support for this hypothesis was much weaker than the support for H2, that hate and dislike differ in degree. Study 2 did not provide the evidence initially expected in regard to Hypotheses 1 and 3 but the results did confirm the same patterns anticipated from Study 1. H1 was partially supported, in that evil of the target was predicted by perceptions of intentionality of the target's actions. However, contrary to H1 (but similar to Study 1), responsibility of the target was not a significant predictor of perceptions of target evil. H2 was supported by several main effects. It was revealed that experiencing hate is more wrong, more resisted, and more likely to occur in response to target evil than dislike. H3 was supported by hate experiences being characterized by more intense feelings of hate than dislike experiences, while intensity of dislike was similar for both experience types. Hate and dislike intensity also differed by subject, in that the self experienced similar intensity of dislike to others, but less intense hate. Additionally, hate experiences were characterized by a stronger action tendency than dislike experiences.

**Exploratory Analyses.** There was one additional hypothesis that was not adequately addressed by the planned analyses. Hypothesis 4 is that there are two distinct types of hate, one that is justified and one that is not. These distinctions were predicted to include the following: (a) hate toward groups is more wrong than hate toward individuals (particularly when experienced by others), (b) that others' hate is morally worse than the self's hate, and (c) that the target of the self's hate is more evil than the target of others'. In the planned analyses of Study 2, hate was

compared to dislike, so it was not possible to look for distinctions in types of hate. Therefore, some of the analyses from Study 2 were repeated, including only the participants who answered questions about hate. These analyses include 390 participants that did not differ significantly in gender, race, or age from the excluded participants.

H4(a) is that hate toward groups is more wrong than hate toward individuals (particularly when experienced by others). A 2 (subject: self, other)  $\times$  2 (target: group, individual) analysis of variance was conducted on the variable measuring how wrong it was to experience hate. As predicted, a main effect of target revealed that it was considered more wrong to hate groups ( $M = 3.84$ ,  $SE = .14$ ) than to hate individuals ( $M = 3.38$ ,  $SE = .14$ ),  $F(1,387) = 5.15$ ,  $p = .02$ ,  $\eta_p^2 = .01$ . Furthermore, there was a two-way interaction between subject and target,  $F(1,387) = 5.55$ ,  $p = .01$ ,  $\eta_p^2 = .01$ . Subsequent simple effects tests with a Bonferroni correction revealed that others hate toward groups ( $M = 4.66$ ,  $SE = .29$ ) was worse than others' hate toward individuals ( $M = 3.73$ ,  $SE = .20$ ),  $F(1,387) = 9.80$ ,  $p = .002$ ,  $\eta_p^2 = .03$ , or the self's hate toward groups ( $M = 3.02$ ,  $SE = .20$ ),  $F(1,387) = 31.47$ ,  $p < .001$ ,  $\eta_p^2 = .08$ . There was no difference in hate experienced by the self toward groups ( $M = 3.02$ ,  $SE = .20$ ) or individuals ( $M = 3.04$ ,  $SE = .19$ ).

The second part of H4 is that others' hate is morally worse and less justified than the self's hate. Support for H4(b) would include main effects of subject for the previously described 2 (subject: self, other)  $\times$  2 (target: group, individual) analysis of variance conducted on the variable measuring how wrong it was to experience hate, as well as the analysis of how much the subject resisted. A main effect of subject on wrongness revealed that others ( $M = 4.19$ ,  $SE = .15$ ) were significantly more wrong for experiencing hate than was the self ( $M = 3.03$ ,  $SE = .14$ ),  $F(1,387) = 33.21$ ,  $p < .001$ ,  $\eta_p^2 = .08$ . There was no main effect of subject or target on resistance to hate.

Relatedly, H4(c) is that the target of the self's hate is worse than the target of others' hate. As predicted, a main effect of subject revealed that the target of the self's hate ( $M = 4.80$ ,  $SE = .13$ ) was more evil than the target of others' hate ( $M = 3.92$ ,  $SE = .14$ ),  $F(1,387) = 20.70$ ,  $p < .001$ ,  $\eta_p^2 = .05$ .

The results of the exploratory analyses support H4 and indicate that there are differences in hate. Hate toward groups is more wrong than hate toward individuals, especially when experienced by others. Furthermore, hate that others feel is less justified than hate felt by the self because it is more wrong. Finally, the target of hate for the self is more evil than the target of others' hate. With the previously reported main effects of emotion from the planned analyses in which hate was both more wrong and more resisted than dislike, these results indicate that people justify their own hatred more than other people's. The target of the emotion for the self is perceived to be worse, so naturally hating them is more justified than other people's hate, which is considered more wrong. Additionally, people report that they resist negative emotions more than other people do. This seems to suggest a self-serving bias: the target of my emotions is more evil, so my negative emotions are justified, but even so I resist it more than you do.

### **General Discussion**

Though Studies 1 and 2 differed in their methodologies, and to some extent the hypotheses they tested, both were designed to identify the determinants of evil and examine the relationship between evil, hate, and dislike. Hypotheses 1-3 were tested across both studies, while Hypotheses 4 was only tested in Study 2.

Hypothesis 1 was that the necessary determinants by which people identify evil in others are intentionality of the target and responsibility of the target for his or her actions. Both Studies 1 and 2 revealed that intentionality of the target, but not responsibility of the target is a

determinant of the target's evil, though neither study provided evidence that both are necessary for the determination of evil. The test of this in Study 1 was not executed as well as it should have been. The stories either did not sufficiently manipulate responsibility of the target, or people did not understand the intended concept of responsibility of the target (either because of language or because of a failure to differentiate responsibility from intentionality), or both. Regardless, Study 2 also indicated that responsibility of the target did not predict the perceived evil of the target.

Hypothesis 2, that hate and dislike differ in degree, was supported by both studies. Based on my previous experiences studying hate, it was predicted that anticipated feelings of hate would be weaker in Study 1 than anticipated feelings of dislike, but that in recounting personal experiences, both hate and dislike intensity would be higher in hate experiences compared with dislike experiences. These patterns were supported for both Studies 1 and 2. Study 2 revealed that hate and dislike differed in degree in perceptions of evil of the target, resistance to feeling the emotion, wrongness of feeling the emotion, and that hate had a stronger action tendency than dislike.

Hypothesis 3 was that hate and dislike differ in kind. Regarding kind, there were few differences in Study 1 between hate and dislike, but importantly, Study 1 only measured participants' anticipated feelings of hate and dislike, were they the victims or witnesses of wrongdoing. It was not a personal situation. Of note, though, is that in both Studies 1 and 2 feelings of hate and dislike independently predicted participants' perceptions that the target was evil. Study 2 revealed more about the differences of hate and dislike in kind, by comparing experiences of hate and dislike to one another. The target of hate was considered more evil than

the target of dislike. It was also considered more wrong to feel hate, and hate was more resisted than dislike.

Hypothesis 4 was that there are distinct types of hate. This hypothesis was supported by evidence that hate toward groups reveals more evil about the subject than hate toward individuals, especially when experienced by others. Additionally, others were considered more wrong in experiencing hate than the self, and others were less likely to resist feeling hate than the self. Finally, a target was considered more evil when they offended the self relative to a target who has offended someone else.

Overall, the goal of these studies was to broaden our understanding of evil and hate by identifying the determinants of perceptions of evil in others, confirming the relationship between hate and dislike, and defining the complex relationship between hate and evil. In regard to the identification of the determinants of evil, despite the failed manipulation (or failure of participants to understand the question which assessed the manipulation), overall perceptions of intentionality and responsibility were both statistically significant determinants of evil in Study 1. This analysis along with the failure of the manipulation question suggests that there might be a dimension related to responsibility that people recognize, but do not label with the term *responsibility*. This could be attributed to variation in how people interpret the meaning of the word responsibility, or an indication that there is another dimension that is important to perceptions of evil. Further research is needed to explore these dimensions more thoroughly.

Contrary to predictions based on previous research (Van Bavel et al., 2018), hate and dislike differed much more in degree than in kind. While there were slight differences in hate and dislike in kind on items like resistance, wrongness of feeling, and evil of target, hate seemed to generally be worse in degree than dislike. Hate indicated more evil in the target, was more

wrong for the subject to feel, and was more resisted by the subject. However, the aforementioned dependent variables of interest were all moral in nature in Study 2 and in Study 1 the anticipated feelings of hate and dislike were in response to immoral acts. Therefore, the degree to which hate is worse than dislike appears to be a moral one. Additionally, in both studies hate and dislike independently predicted evil of the target, which indicates that they are, at least in part, differentially related to evil.

Though hate and dislike did not differ as much in kind as was predicted on the moral outcome variables (wrongness of feeling, resistance, and evil of target), they did differ on action tendency. This outcome might provide an explanation for the failure of the moral dimensions and an important distinction in the nature of hate and dislike. Action tendency is an important distinction between attitudes and emotions. Attitudes can be thought of as “enduring affective valuations”, whereas emotions are “occurrent affective reactions that mobilize relational behavior” (Gervais & Fessler, 2017, p. 3). In other words, attitudes are general evaluations while emotions are momentary reactions that *incite action toward the object*. In previous research, we demonstrated that hate behaves as an emotion in that it is brief and incites action (Merrick & Beike, 2019). Dislike has also been categorized as an emotion (Roseman, 1984), though it is more often thought of as an attitude (Collins, 1970; Petty, Ostrom, & Brock, 2014; Zimbardo, Ebbesen, & Maslach, 1977). Attitudes have even been defined in terms of like and dislike, “attitudes are the core of our likes and dislikes for certain people, groups, situations, objects, and intangible ideas” (Zimbardo, Ebbesen, & Maslach, 1977, p.20). The current research demonstrated that hate and dislike differ in their action tendencies, such that hate is more likely to incite action than dislike supporting the idea that hate is an emotion while dislike is an attitude. That hate and dislike do not differ on moral determinants suggests that hate and dislike

may be part of the same higher affective system or sentiment, but the evidence that they differ in degree of action tendency suggests that they differ in kind of construct.

Study 2 demonstrated a complex relationship between hate and evil. Several key findings revealed that value judgments are made about the target and the subject of hate that serve to justify (or deem unjustified) the emotion, and therefore determine which (the subject or target) is more evil. First, hating groups was revealed to be more wrong than hating individuals. This was expected, as hate toward groups often denotes prejudice and the more aggressive intergroup hate (Sternberg, 2003). Interestingly, though, groups were considered more evil than individuals. By itself, it is unsurprising that groups were considered more evil than individuals. This finding suggests that people are receptive to the concept of deindividuation, or the loss of self and morality associated with being subsumed by a group (Reicher, Spears, Postmes, 1995). Groups have more power and more influence, and therefore it makes sense that they are considered more evil. However, taken together with the previous finding that hating groups is more wrong than hating individuals reveals an interesting paradox. Groups are more evil than individuals, and yet it is more evil to feel hate toward groups. This conundrum highlights the original research question, if hate reveals what is evil, how can it be wrong to feel hate? It seems that both are possible – hate can reveal what is evil, but that doesn't make it justified. The key to this conundrum is the two-way interaction that was observed. It is wrong to hate groups, *but especially when someone else hate groups*. An important distinction is made between the self and other. Hate is safer in my hands than in yours.

Relatedly, Study 2 also revealed that others' hate is more wrong than the self's hate. When combined with the findings that the target of the self's hate was considered more evil than the target of others' hate, it once again suggests that the self's hate is more justified than

someone else's. The justification for the self's hate has thus far only been applied to hate toward groups but might also be considered for individuals. Research has demonstrated that stereotypes are used as justification for prejudice toward groups – even in the absence of specific violations (Crandall, Bahns, Warner, & Schaller, 2011). There are two important ways in which these findings support the current research. First, if evil is used as the basis of a stereotype, then hating someone or some group might feel justified based on that categorization. This would apply to both individuals and groups. Second, is the justification itself. People know that prejudice is wrong, so they feel the need to justify it and use stereotypes to do so. In the case of hate, people believe that feeling hate is wrong, so they feel the need to justify the feeling.

This explanation encompasses the justification of the self's hate. However, the present research also demonstrates that people do not justify others' hate. That people justify their own hate but not the hate of others suggests an attribution error, or at least a difference in how people perceive their own hate relative to someone else's (Jones & Nisbett, 1987). How positively the self is viewed is a key factor in self-evaluation (Sedikides & Strube, 1997). If people generally categorize hate as evil (Halperin, 2008), but then feel hate, they are motivated to evaluate the hate as justified by deeming the target to be evil. This enables them to avoid negatively evaluating the self. On the other hand, they are not motivated to do this for others, and so their hate is considered more evil, their target less evil, and therefore their hate is less justified.

Together, Study 2 provides evidence that people use dissonance reduction to justify their hate. People have a notion that it is wrong to experience hate so when they do, they justify the hate by relying on character assumptions about the target. When others experience hate, it must be a reflection of the character of the subject, rather than of the target. The paradox between self and other indicates that the most important criterion for defining the two types of hate might not

be whether the hate is toward a group or individual (though there are differences), or whether the hate is justified or unjustified. Rather, we might make assumptions about justified and unjustified hate based on our attempts to reduce dissonance. The paradox between self and other also indicates something about the balance of evil in hate. It suggests a revolving attribution of evil – the evil shifts from the subject to the target depending on the justification for the hate. For the self, hate is a necessary evil to combat evil of the target. For the other, it is giving in to evil within, since the target is not evil enough to justify it.

### **Limitations**

One limitation of this research is the questions that were meant to differentiate intentionality and responsibility. In Study 1, participants seemed to understand that there were differences in responsible and not responsible vignettes regarding evil, though the questions that checked the manipulation apparently failed. Therefore, it is possible that the concept itself was successful, but the language of the question was ineffective at capturing the difference between responsibility and intentionality. Future research could address this limitation by administering the same vignettes to participants but asking them different questions about the responsibility and intentionality of the Focal Character (e.g. “How blameworthy was the Focal Character?”), to see if it is the concept or the question that is ineffective.

A limitation of Study 2 is that there are likely confounds between the types of groups and individuals that participants reported feeling emotions toward. For instance, someone who hates a political figure (e.g., Donald Trump) might be using that person to symbolically represent an entire group (e.g., Republicans). This confound could be addressed by looking more specifically at the individuals and groups that participants reported feeling dislike or hate toward and coding for group/person type, and whether the group is achieved or ascribed (Allport, 1954). A similar

confound is possible in the self/other condition in Study 2. Based on the attribution bias, it is likely that people are more apt to disclose an unacceptable target for others than for the self (e.g., someone *else* hates Black people. I hate racists) (Jones & Nisbett, 1987). Therefore, the social acceptability of hatred toward the group should also be coded. This coding would not be as effective for coding responses toward individuals, as those answers are generally someone the participant knows, and therefore would be impossible to decipher without further information.

### **Future Directions**

The current set of studies provides new information about evil and hate which sets the foundation for new studies. Study 2 provided some evidence that there are different types of hate, or at least that people have different perceptions about hate, based on who is feeling it and the target toward which it is felt. Additional studies should look more closely at the distinction between these hates by examining their motivations and their consequences. An extension of this particular study would benefit from additional questions that capture the ideas of justification of hate and whether that differs based on the subject or target of hate. Another future direction from this set of studies comes from the conclusion that hate and dislike differ in degree much more than they differ in kind. If so, it would be interesting to know whether the difference of degree in hate and dislike is based on degrees of evil.

### **Conclusions**

This pair of studies examined the determinants of evil and the relationship between evil, hate, and dislike. Together, they demonstrate that evil is perceived in actions that are intentional. They also demonstrate that hate and dislike differ much more in degree than in kind, but that their difference of kind is based on their relationship with evil. Furthermore, a closer look at hate and evil revealed a complex relationship. the relationship between hate and evil is

unquestionably complex. Hate informs us about what is evil more than dislike, so it must serve a purpose, and yet hate is considered more wrong than dislike. The target of my hate is more evil than the target of other people's, and yet I resist hate more than they do. The original research question posed here was this: if hate informs us about what is evil, then how can it be wrong to feel hate? Rather than answer this question, however, this research has changed it from a question to a paradox. Hate informs us about what is evil, and yet it is wrong to feel hate.

## References

- Allport, G.W. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.
- Aumer-Ryan, K., & Hatfield, E. C. (2007). The design of everyday hate: A qualitative and quantitative analysis. *Interpersona: An International Journal on Personal Relationships*, 1(2), 143-172. <https://dx.doi.org/10.5964/ijpr.v1i2.11>
- Baumeister, R. F. (1997). *Evil: Inside human cruelty and violence*. New York: Freeman.
- Baumeister, R. F., & Butz, D. A. (2005). Roots of hate, violence, and evil. In R. J. Sternberg (Ed.), *The psychology of hate* (pp. 87–102). Washington, DC: American Psychological Association. <http://dx.doi.org/10.1037/10930-005>
- Beck, A. T. (1999). *Prisoners of hate: The cognitive basis of anger, hostility, and violence*. New York: Harper Collins.
- Ben-Ze'ev, A. (1992) Anger and hate. *Journal of Social Philosophy*, 23(2), 85-110. <https://dx.doi.org/10.1111/j.1467-9833.1992.tb00295.x>
- Ben-Ze'ev, A. (2000) *The subtlety of emotions*. Cambridge, MA: MIT Press
- Book, A.S., Visser, B., & Volk, A. (2014). Unpacking evil: claiming the core of the dark triad. *Personality and Individual Differences*, 73, 29–38. <http://dx.doi.org/10.1016/j.paid.2014.09.016>.
- Book, A., Visser, B. A., Blais, J., Hosker-Field, A., Methot-Jones, T., Gauthier, N. Y., ... D'Agata, M. T. (2016). Unpacking more “evil”: What is at the core of the dark tetrad? *Personality and Individual Differences*, 90, 269–272. <https://dx.doi.org/10.1016/j.paid.2015.11.009>
- Brontë, E. (2003). *Wuthering heights*. London; New York: Penguin Books (Original work published 1847)
- Brumbach, B. H., Figueredo, A. J., & Ellis, B. J. (2009). Effects of harsh and unpredictable environments in adolescence on development of life history strategies: A longitudinal test of an evolutionary model. *Human Nature*, 20(1), 25–51. <https://dx.doi.org/10.1007/s12110-009-9059-3>
- Campbell, J., Schermer, J. A., Villani, V. C., Nguyen, B., Vickers, L., & Vernon, P. A. (2009). A behavioral genetic study of the dark triad of personality and moral development. *Twin Research and Human Genetics*, 12(2), 132–136. <https://dx.doi.org/10.1375/twin.12.2.132>
- Collins, B.E. (1970) *Social psychology*. Reading, Massachusetts: Addison-Wesley, 1970.

- Crandall, C.S., Bahns, A.J., Warner, R., Schaller, M. (2011) Stereotypes as justification of prejudice. *Personality and Social Psychology Bulletin*, 37(11), 1488-1498. <http://dx.doi.org/10.1177/0146167211411723>
- Darley, J. M. (1992). Social organization for the production of evil. *Psychological Inquiry* 3, 199-218. <https://dx.doi.org/45.26.86.55>
- Darley, J. M., & Batson, C. D. (1973). "From Jerusalem to Jericho": A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology*, 27(1), 100–108. <https://dx.doi.org/10.1037/h0034449>
- Darley, J. M., & Latane, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, 8(4, Pt.1), 377–383. <https://dx.doi.org/10.1037/h0025589>
- Darwin, C. (1872). *The expression of the emotions in man and animals*. London, UK: Murray.
- Dozier, R. W., Jr. (2002). *Why we hate: Understanding, curbing, and eliminating hate from ourselves and our world*. Chicago: Contemporary Books.
- Ekman, P. (2016). What scientists who study emotion agree about. *Perspectives on Psychological Science*, 11(1), 31–34. <https://dx.doi.org/10.1177/1745691615596992>
- Ellsworth, P. C. (2013). Appraisal theory: Old and new questions. *Emotion Review*, 5(2), 125–131. <https://dx.doi.org/10.1177/1754073912463617>
- Elster, J. (1999). *Alchemies of the mind: Rationality and the emotions*. Cambridge: Cambridge University Press.
- Fischer, A., Halperin, E., Canetti, D., & Jasini, A. (2018). Why we hate. *Emotion Review*. <https://dx.doi.org/10.1177/1754073917751229>
- Fitness, J. (2000). Anger in the workplace: an emotion script approach to anger episodes between workers and their superiors, co-workers and subordinates. *Journal of Organizational Behavior*, 21(2), 147-162. [https://dx.doi.org/10.1002/\(SICI\)1099-1379\(200003\)21:2<147::AID-JOB35>3.0.CO;2-T](https://dx.doi.org/10.1002/(SICI)1099-1379(200003)21:2<147::AID-JOB35>3.0.CO;2-T)
- Fitness, J., and Fletcher, G.J. (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://dx.doi.org/10.1037/0022-3514.65.5.942>
- Frijda, N.H. (1986). *The emotions*. Cambridge: Cambridge University Press.
- Frijda, N. H. (1993). The place of appraisal in emotion. *Cognition and Emotion*, 7(3–4), 357–387. <https://dx.doi.org/10.1080/02699939308409193>

- Fromm, E. (1973). *Anatomy of human destructiveness*. New York: Holt.
- Furnham, A., Richards, S., & Paulhus, D. (2013). The dark triad of personality: A ten year review. *Social and Personality Psychology Compass*, 7, 199–216. <http://dx.doi.org/10.1111/spc3.12018>.
- Gaylin, W. (2003). *Hatred: The psychological descent into violence*. New York: Public Affairs.
- Gervais, M. M. & Fessler, D. M. (2017). On the deep structure of social affect: Attitudes, emotions, sentiments, and the case of “contempt”. *Behavioral and Brain Sciences*, 40 (e225), 1-18. <https://dx.doi.org/10.1017/S0140525X16000352>.
- Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, 101(2), 366–385. <https://dx.doi.org/10.1037/a0021847>
- Graham, J., & Haidt, J. (2012). Sacred values and evil adversaries: A moral foundations approach. In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil*. (pp. 11–31). Washington, DC: American Psychological Association <http://dx.doi.org/10.1037/13091-001>
- Graham, J., Haidt, J., & Nosek, B.A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, 96(5), 1029–1046. <https://dx.doi.org/10.1037/a0015141>
- Gray, K., & Wegner, D. M. (2011). Dimensions of moral emotions. *Emotion Review*, 3(3), 258–260. <https://dx.doi.org/10.1177/1754073911402388>
- Haidt, J. (2003). The moral emotions. In R. J. Davidson, K. R. Scherer & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 852–870). Oxford, UK: Oxford University Press.
- Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus, Fall*, 330(4), 55-66. <https://dx.doi.org/10.1162/0011526042365555>
- Halperin, E. (2008). Group-based hatred in intractable conflict in Israel. *Journal of Conflict Resolution*, 52(5), 713-736. <https://dx.doi.org/10.1177/0022002708314665>
- Halperin, E., Canetti, D., & Kimhi, S. (2012). In love with hatred: Rethinking the role hatred plays in political behavior. *Journal of Applied Social Psychology*, 42(9), 2231–2256. <https://dx.doi.org/10.1111/j.1559-1816.2012.00938.x>
- Harms, P. D., Spain, S., & Hannah, S. (2011). Leader development and the dark side of personality. *Leadership Quarterly*, 22(3), 495–509. <https://dx.doi.org/10.1016/j.leaqua.2011.04.007>

- Hodson, G., Hogg, S. M., & MacInnis, C. C. (2009). The role of “dark personalities” (narcissism, Machiavellianism, psychopathy), Big Five personality factors, and ideology in explaining prejudice. *Journal of Research in Personality*, 43(4), 686–690. <https://dx.doi.org/10.1016/j.jrp.2009.02.005>
- Horowitz, D. L. (2001). *The deadly ethnic riot*. Berkeley: University of California Press.
- James, W. (1890) *The principles of psychology*. New York: Holt.
- Johnson-Laird, P. N., & Oatley, K. (1992). Basic emotions, rationality, and folk theory. *Cognition and Emotion*, 6(3–4), 201–223. <https://dx.doi.org/10.1080/02699939208411069>
- Jonason, P. K., Zeigler-Hill, V., & Okan, C. (2017). Good v evil: Predicting sinning with dark personality traits and moral foundations. *Personality and Individual Differences*, 104, 180–185. <https://dx.doi.org/10.1016/j.paid.2016.08.002>
- Jones, E. E., & Nisbett, R. E. (1987). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 79-94). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.
- Kant, I. (2002). *Groundwork for the metaphysics of morals* (A. W. Wood, Ed. & Trans.). New Haven, CT: Yale University Press. (Original work published 1785)
- Lewis, M. (1993). Self-conscious emotions: Embarrassment, pride, shame, and guilt. In M. Lewis & J. Haviland (Eds.), *Handbook of emotions* (pp. 563-573). New York: Guilford Press.
- Matsumoto, D., Hwang, H. C., & Frank, M. G. (2017). Emotion and aggressive intergroup cognitions: The ANCODI hypothesis. *Aggressive Behavior*, 43(1), 93-107. <https://dx.doi.org/10.1002/ab.21666>
- McCauley, C. (2002). Psychological issues in understanding terrorism and the response to terrorism. In C. Stout (Ed.), *The psychology of terrorism: Theoretical understandings and perspectives, Vol. III*, (pp. 3-30). Westport, CT: Praeger Publishers.
- Merrick C.R. & Beike D.R. (2019) *Appraisals, Action Tendencies, and Relational Aspects of the Emotion of Interpersonal Hate*. Submitted for publication.
- Milgram, S. (1974). *Obedience to authority*. New York: Harper & Row.
- Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal theories of emotion: State of the art and future development. *Emotion Review*, 5(2), 119-124. <https://dx.doi.org/10.1177/1754073912468165>

- Oatley, K., & Johnson-Laird, P. N. (1987). Towards a cognitive theory of emotions. *Cognition and Emotion*, 1(1), 29–50. <https://dx.doi.org/10.1080/02699938708408362>
- Opatow, S. (1990). Moral exclusion and injustice: An introduction. *Journal of Social Issues* 46(1), 1–20. <https://dx.doi.org/10.1111/j.1540-4560.1990.tb00268.x>
- Opatow, S. (2004). Hate, conflict, and moral exclusion. In R. Sternberg (Ed.), *The psychology of hate*. Washington, D.C.: APA Books. <http://dx.doi.org/10.1037/10930-007>
- Paulhus, D. L., & Williams, K. M. (2002). The Dark Triad of personality: Narcissism, Machiavellianism and psychopathy. *Journal of Research in Personality*, 36(6), 556–563. [https://dx.doi.org/10.1016/S0092-6566\(02\)00505-6](https://dx.doi.org/10.1016/S0092-6566(02)00505-6)
- Pearson, A. R., Dovidio, J. F., & Pratto, F. (2007). Racial prejudice, intergroup hate, and blatant and subtle bias of whites toward blacks in legal decision making in the United States. *International Journal of Psychology & Psychological Therapy*, 7(2), 145–158. <https://www.redalyc.org/pdf/560/56070203.pdf>
- Petty, R., Ostrom, T.M., Brock, T.C. (2014) *Cognitive responses in persuasion*. New York: Psychology Press.
- Reicher, S., Levine, R. M., & Gordijn, E. (1998). More on deindividuation, power relations between groups and the expression of social identity: Three studies on the effects of visibility to the in-group. *British Journal of Social Psychology*, 37(1), 15–40. <https://dx.doi.org/10.1111/j.2044-8309.1998.tb01155.x>
- Roseman, I. J. (1984). Cognitive determinants of emotion: A structural theory. In P. Shaver (Ed.), *Review of personality and social psychology*, Vol. 5: *Emotions, relationships, and health* (pp. 11–36). Beverly Hills, CA: SAGE.
- Rozman, E.B., McCauley, C., and Rosin, P. (2005). From Plato to Putnam: Four ways to think about hate. In R.J. Sternberg (Ed.), *The psychology of hate* (pp. 3-36). Washington, DC: American Psychological Association. <https://dx.doi.org/10.1037/10930-001>
- Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology*, 76(4), 574–586. <http://dx.doi.org/10.1037/0022-3514.76.4.574>
- Scherer, K. R. (1984). On the nature and function of emotion: A component process approach. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 293–318). Hillsdale, NJ: Lawrence Erlbaum.

- Sedikides, C., Strube, M.J. (1997) Self evaluation: To thine own self be good, to thine own self be sure, to thine own self be true, and to thine own self be better. In Zanna, M.P. (Ed.), *Advances in Experimental and Social Psychology* (pp. 209-9269). San Diego, CA: Academic Press.
- Shand, A. F. (1920). *The foundations of character: Being a study of the tendencies of the emotions and sentiments*. London: Macmillan and Company, Limited.
- Shweder, R. A., Much, N. C, Mahapatra, M., & Park, L. (1997). The "Big Three" of morality (autonomy, community, divinity) and the "Big Three" explanations of suffering. In A. Brandt & P. Rozin (Eds.), *Morality and health* (pp. 119-169). New York: Routledge.
- Singer, M. G. (2004). The concept of evil. *Philosophy*, 79(2), 185-214. <https://dx.doi.org/98.185.174.101>
- Smith, E. R., & Mackie, D. M. (2015). Dynamics of group-based emotions: Insights from intergroup emotions theory. *Emotion Review*, 7(4), 349–354. <https://dx.doi.org/10.1177/1754073915590614>
- Staub, E. (1999). The roots of evil: Social conditions, culture, personality, and basic human needs. *Personality and Social Psychology Review*, 3(3), 179–192. [https://dx.doi.org/10.1207/s15327957pspr0303\\_2](https://dx.doi.org/10.1207/s15327957pspr0303_2)
- 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). Understanding and combating hate. In R. J. Stern- berg (Ed.), *The psychology of hate* (pp. 37–49). Washington, DC: American Psychological Association. <http://dx.doi.org/10.1037/10930-002>
- Tangney, J. P. (1999). The self- conscious emotions: Shame, guilt, embarrassment and pride. *Handbook of cognition and emotion*, 541-568. <https://dx.doi.org/10.1002/0470013494.ch26>
- Van Bavel, J. J., Ray, J. L., Granot, Y., & Cunningham, W. A. (2018). *The psychology of hate: Moral concerns differentiate hate from dislike*. Unpublished manuscript. <https://dx.doi.org/10.31234/osf.io/x9y2p>
- Weiner, B., (1993). On sin versus sickness: A theory of perceived responsibility and social motivation. *American Psychologist*, 48(9), 957-965. <http://dx.doi.org/10.1037/0003-066X.48.9.957>
- Zimbardo, P. (2006). A situationist perspective on the psychology of evil: Understanding how good people are transformed into perpetrators. In R. Falk, I. Gendzier, & R. J. Lifton (Eds.), *Crimes of War: Iraq*. (pp. 366–369). New York, NY: Nation Books.

- Zimbardo, P. G. (1975). On transforming experimental research into advocacy for social change. In M. Deutsch & H. Hornstein (Eds.), *Applying social psychology: Implications for research, practice, and training* (pp. 33-66). Hillsdale, NJ: Erlbaum.
- Zimbardo, P.G., Ebbesen, E.B., & Maslach, C. (1977). *Techniques of attitude measurement: Influencing attitudes and changing behavior*. Menlo Park, CA: Addison-Wesley Publishing.
- Zimbardo, P. G., Haney, C., Banks, C., & Jaffe, D. (1973, April 8). The mind is a formidable jailer: A Pirandellian prison. *The New York Times Magazine*, pp. 38.

**Tables**

Table 1  
*Study 1 Fixed Effects ANOVA for Perceived Evil of Focal Character  
 (Full Sample)*

Source	<i>df</i>	Sum Square	Mean Square	F	p	$n_p^2$
Between						
Intention	1	1585.83	1585.83	262.24	<.001	.40
Responsible	1	72.84	72.84	12.05	.001	.03
Intention*Responsible	1	.61	.61	.10	.75	<.001
Within						
Violation	5	1165.89	233.18	126.87	<.001	.24
Violation*Intention	5	549.98	110.00	59.85	<.001	.13
Violation*Responsible	5	63.22	12.64	6.88	<.001	.02
Violation*Intention* Responsible	5	11.69	2.34	1.27	.27	.003

\*\*p < 0.01

Table 2  
*Study 1 Repeated Measures ANOVA for Anticipated Hate and Dislike toward Focal Character (Full Sample)*

Source	<i>df</i>	Sum Square	Mean Square	F	p	$\eta_p^2$
Between						
Intention	1	2451.96	2451.96	160.00	<.001	.30
Responsible	1	4.18	4.18	.273	.602	.001
Intention*Responsible	1	15.15	15.15	.99	.32	.003
Within						
Violation	5	1041.08	208.22	59.92	<.001	.14
Violation*Intention	5	1398.42	279.68	80.49	<.001	.17
Violation*Responsible	5	64.34	12.87	3.70	.002	.01
Violation*Responsible*Intention	5	45.22	9.04	2.60	.02	.007
Emotion	1	1073.33	1073.33	290.46	<.001	.43
Emotion*Responsible	1	.15	.15	.04	.84	<.001
Emotion*Intention	1	10.37	10.37	.281	.10	.007
Emotion*Responsible*Intention	1	5.78	5.78	1.57	.21	.004
Violation*Emotion	5	13.54	2.71	3.62	.003	.009
Violation*Emotion*Responsible	5	1.32	.26	.35	.88	.001
Violation*Emotion*Intention	5	11.65	2.33	3.12	.008	.008
Violation*Emotion*Responsible*Intention	5	4.03	.81	1.08	.37	.003

\*\*p < 0.01

Table 3  
*Study 1 Fixed Effects ANOVA for Perceived Evil of Focal Character  
 (Sample with Exclusion Criterion Applied)*

Source	<i>df</i>	Sum Square	Mean Square	F	p	$\eta_p^2$
Between						
Intention	1	167.27	167.27	29.47	<.001	.10
Responsible	1	5.64	5.64	.99	.32	.004
Intention*Responsible	1	<.001	<.001	<.001	.99	<.001
Within						
Violation	5	113.56	22.71	12.43	<.001	.05
Violation*Responsible	5	31.32	6.27	3.43	.004	.01
*Violation*Intention	5	24.20	4.84	2.65	.02	.01
*Violation*Responsible *Intention	5	15.49	3.10	1.7	.13	.007

\*\*p<.01

Table 4  
*Study 1 Repeated Measures ANOVA for Anticipated Hate and Dislike  
 (Sample with Exclusion Criterion Applied)*

Source	<i>df</i>	Sum Square	Mean Square	F	p	$\eta_p^2$
Between						
Intention	1	221.86	221.86	15.22	<.001	.06
Responsible	1	2.40	2.40	.17	.685	.001
Intention*Responsible	1	.005	.005	<.001	.99	<.001
Within						
Violation	5	94.55	18.91	5.10	<.001	.02
Violation*Intention	5	116.18	23.24	6.27	<.001	.03
Violation*Responsible	5	23.12	4.62	1.25	.29	.005
Violation*Responsible*Intention	5	14.89	2.97	.80	.55	.003
Emotion	1	59.43	59.43	18.18	<.001	.08
Emotion*Responsible	1	5.08	5.08	2.76	.10	.01
Emotion*Intention	1	1.35	1.35	.74	.39	.003
Emotion*Responsible*Intention	1	5.50	5.50	2.99	.09	.01
Violation*Emotion	5	37.35	7.47	6.75	<.001	.03
Violation*Emotion*Responsible	5	9.12	1.82	1.65	.15	.007
Violation*Emotion*Intention	5	7.55	1.51	1.36	.24	.005
Violation*Emotion*Responsible*Intention	5	11.79	2.36	2.13	.06	.009

\*\*p < 0.01

Table 5  
*Study 2 Fixed Effects ANOVA for Evil of Target*  
*(Full Sample)*

<b>Source</b>	<b><i>df</i></b>	<b>Sum Square</b>	<b>Mean Square</b>	<b>F</b>	<b>p</b>	<b><math>\eta_p^2</math></b>
Subject	1	99.85	99.85	25.40	<.001	.03
Target	1	33.23	33.23	8.45	.004	.01
Emotion	1	96.87	96.87	24.64	<.001	.03
Subject*Target	1	1.04	1.04	.26	.61	<.001
Subject*Emotion	1	5.78	5.78	1.47	.23	.002
Target*Emotion	1	3.05	3.05	.78	.38	.001
Subject*Target* Emotion	1	4.41	4.41	1.12	.29	.001

\*\*p < 0.01

Table 6  
*Study 2 Fixed Effects ANOVA for Wrongness of Emotion*  
*(Full Sample)*

<b>Source</b>	<b><i>df</i></b>	<b>Sum Square</b>	<b>Mean Square</b>	<b>F</b>	<b>p</b>	<b><math>\eta_p^2</math></b>
Subject	1	248.87	248.87	70.08	<.001	.08
Target	1	88.99	88.99	25.06	<.001	.03
Emotion	1	75.76	75.76	21.33	<.001	.03
Subject*Target	1	46.30	46.30	13.04	<.001	.02
Subject*Emotion	1	.42	.42	.13	.72	<.001
Target*Emotion	1	.51	.51	.15	.70	<.001
Subject*Target* Emotion	1	4.79	4.79	1.42	.23	.002

\*\*p < 0.01

Table 7  
*Study 2 Fixed Effects ANOVA for Resistance of Emotion*  
*(Full Sample)*

<b>Source</b>	<b><i>df</i></b>	<b>Sum Square</b>	<b>Mean Square</b>	<b>F</b>	<b>p</b>	<b><math>\eta_p^2</math></b>
Subject	1	24.72	24.72	9.73	.002	.01
Target	1	10.88	10.88	4.28	.04	.005
Emotion	1	38.67	38.67	15.22	<.001	.02
Subject*Target	1	.36	.36	.14	.71	<.001
Subject*Emotion	1	1.12	1.12	.44	.51	.001
Target*Emotion	1	.05	.05	.02	.89	<.001
Subject*Target* Emotion	1	10.98	10.98	4.32	.04	.006

\*\*p < 0.01

Table 8  
*Study 2 Repeated Measures ANOVA for Hate and Dislike Intensity  
 (Full Sample)*

Source	<i>df</i>	Sum Square	Mean Square	F	p	$\eta_p^2$
Between						
Subject	1	.18	.18	.05	.82	<.001
Emotion	1	142.41	142.41	40.02	<.001	.05
Target	1	4.59	4.59	1.29	.26	.002
Subject*Emotion	1	8.96	8.96	2.52	.11	.003
Subject*Target	1	.90	.90	.25	.62	<.001
Emotion*Target	1	2.35	2.35	.66	.42	.001
Subject*Emotion* Target	1	.50	.50	.14	.71	<.001
Within						
EmoIntensity	1	471.56	471.56	474.07	<.001	.38
EmoIntensity*Subject	1	8.64	8.64	8.69	.003	.01
EmoIntensity*Emotion	1	52.53	52.53	52.81	<.001	.06
EmoIntensity*Target	1	3.38	3.38	3.40	.07	.004
EmoIntensity*Subject *Emotion	1	1.01	1.01	1.01	.31	.001
EmoIntensity*Subject *Target	1	.30	.30	.30	.58	<.001
EmoIntensity*Emotion* Target	1	2.09	2.09	2.10	.15	.003
EmoIntensity*Subject *Emotion*Target	1	.62	.62	.62	.43	.001

\*\*p < 0.01

Table 9  
*Study 2 Repeated Measures ANOVA for Approach and Avoidance Action Tendency (Full Sample)*

Source	<i>df</i>	Sum Square	Mean Square	F	<i>p</i>	$\eta_p^2$
Between						
Subject	1	479.92	479.92	.73	.39	.001
Emotion	1	4131.12	4131.12	6.27	.01	.008
Target	1	153.87	153.87	.23	.63	<.001
Subject*Emotion	1	457.37	457.37	.69	.41	.001
Subject*Target	1	743.43	743.43	1.13	.29	.001
Emotion*Target	1	576.20	576.20	.88	.35	.001
Subject*Emotion* Target	1	69.07	69.07	.11	.75	<.001
Within						
Action	1	743372.30	743372.30	578.82	<.001	.43
Action*Subject	1	27699.12	27699.12	21.57	<.001	.03
Action*Emotion	1	9119.68	9119.68	7.10	.008	.009
Action*Target	1	4825.75	4825.75	3.76	.06	.005
Action*Subject* Emotion	1	52.72	52.72	.04	.84	<.001
Action*Subject* Target	1	3953.99	3953.99	3.08	.08	.004
Action*Emotion* Target	1	3689.69	3689.69	2.87	.09	.004
Action*Subject* Emotion*Target	1	3667.34	3667.34	2.86	.09	.004

Table 10  
*Study 2 Fixed Effects ANOVA for Evil of Target  
 (Hate Participants Only)*

<b>Source</b>	<b><i>df</i></b>	<b>Sum Square</b>	<b>Mean Square</b>	<b>F</b>	<b>p</b>	<b><math>\eta_p^2</math></b>
Subject	1	76.13	76.13	20.70	<.001	.05
Target	1	27.94	27.94	7.60	.006	.02
Subject*Target	1	4.82	4.82	1.31	.25	.003

\*\*p < 0.01

Table 11  
*Study 2 Fixed Effects ANOVA for Wrongness of Emotion*  
*(Hate Participants Only)*

<b>Source</b>	<b><i>df</i></b>	<b>Sum Square</b>	<b>Mean Square</b>	<b>F</b>	<b>p</b>	<b><math>\eta_p^2</math></b>
Subject	1	131.59	131.59	33.21	<.001	.08
Target	1	20.41	20.41	5.15	.02	.01
Subject*Target	1	21.99	21.99	5.55	.02	.01

\*\*p < 0.01

## Appendices

### Appendix A

#### Study 1 Participants and Results (with exclusion criterion applied)

##### Participants

Participants included 179 University of Arkansas students, who completed the online survey through Qualtrics in exchange for partial fulfillment of a research requirement; and 83 Mechanical Turk workers, recruited through TurkPrime and compensated for their participation based on a ten-dollar hourly rate. The participants included 121 female, 139 male, and two who chose not to answer. They ranged in age from 18 to 67 with a mean age of 25.89. Eighty nine percent of participants were White, 5% were Black, 4% were Asian, 5% were Latinx, less than 1% were Native American, less than 1% were Pacific Islander, and less than 1% indicated other. Participants could choose more than one ethnicity.

##### Results

Hypothesis 1 is that responsibility and intentionality are both necessary for evil to be recognized, but neither is sufficient on its own. A 2 (responsible, not responsible)  $\times$  2 (intentional, unintentional)  $\times$  6 (moral foundations: harm, fairness, ingroup, authority, purity, liberty) mixed model analysis of variance was conducted on the item measuring the perceived evil of the Focal Character (whose actions were manipulated). As predicted, there was a main effect of intention, such that the Focal Characters in vignettes portraying intentional acts ( $M = 3.97$ ,  $SE = .29$ ) were perceived to be more evil than in vignettes portraying unintentional acts ( $M = 2.36$ ,  $SE = .07$ ),  $F(1, 256) = 29.47$ ,  $p < .001$ ,  $\eta_p^2 = .10$ . There was no a main effect of responsibility, though one was predicted.

An additional analysis to explore H1 consisted of computing participants' average scores of intentionality of the Focal Characters' actions, responsibility for the Focal Characters' actions, and perceived evil of the Focal Characters across all vignettes. A simultaneous linear regression was then conducted that demonstrated that intentionality and responsibility were both significant predictors of evil,  $F(2, 257) = 52.70, p < .001$  with an  $R^2 = .29$ . Intentionality was a significant predictor of evil ( $\beta = .32, p = .004$ ). Responsibility was only a marginal predictor of evil ( $\beta = .23, p = .067$ ).

H2 is that hate and dislike differ in degree. A 2 (responsible, not responsible)  $\times$  2 (intentional, unintentional)  $\times$  6 (moral foundations: harm, fairness, ingroup, authority, purity, liberty) mixed model analysis of variance was conducted on the item measuring anticipated feelings of hate and anticipated feelings of dislike. Participants' overall anticipated feelings of hate and dislike toward the Focal Character were different from each other. Anticipated dislike ( $M = 4.28, SE = .16$ ) was more intense overall than anticipated hate ( $M = 3.61, SE = .21$ ),  $F(1, 248) = 18.18, p < .001, \eta_p^2 = .07$ .

H3 is that hate and dislike differ in kind. H3 was explored through the previous analysis that also provided evidence for H2, as well as a simultaneous regression. First, the 2 (responsible, not responsible)  $\times$  2 (intentional, unintentional)  $\times$  6 (moral foundations: harm, fairness, ingroup, authority, purity, liberty) mixed model analysis of variance conducted on the item measuring anticipated feelings of hate and anticipated feelings of dislike was examined. Contrary to predictions, there was no two-way interaction between emotion and responsibility of the target (which is not surprising, given the initial analysis that showed no effect of the responsibility manipulation), nor was there a two-way interaction between emotion and intentionality of the target.

Support for H3 was also sought through a simultaneous linear regression revealed that both anticipated feelings of hate and anticipated feelings of dislike were significant predictors of perceptions of evil,  $F(2, 247) = 79.55, p < .001$  with an  $R^2 = .39$ . As predicted, anticipated hate was a significant predictor of evil ( $\beta = .29, p < .001$ ). Contrary to predictions, however anticipated dislike also predicted evil ( $\beta = .29, p < .001$ ).

## **Appendix B**

Study 1 – each of the vignettes is included for the different conditions. However, the questions for each scenario were the same so the questions will only be presented with the first scenario.

[Participants were randomly assigned to one of four groups: intentional/responsible, intentional/not responsible, unintentional/responsible, unintentional/not responsible. All participants read and responded to six short scenarios. Scenarios were randomized.]

### **Intentional (Responsible):**

#### *Harm*

Benny goes out with to a bar one night with a friend, where his friend starts a fight and encourages Benny to join in. Benny excitedly joins in the fight and starts kicking a stranger while he's on the ground.

Benny hurt the stranger on purpose.

Benny chose to hurt the stranger on his own.

#### *Fairness*

Floyd works for a politician, who doesn't mind lying or fudging the numbers about his accomplishments. During the election, Floyd offers to distribute untrue, slanderous information about the opposing candidate to help his boss win.

Floyd slandered the other candidate on purpose.

Floyd chose to act unfairly on his own.

#### *Ingroup*

Arnold is part of a secret society. He has a friend who is not part of the society, and who asks him to share some secrets about his group that he isn't supposed to tell. Arnold tells his friend the secrets of his group.

Arnold intentionally told his friend secret information about his group.

Arnold chose to share the secret information on his own.

### *Authority*

Jermaine goes with his friend to a rally protesting the government, where people are shouting and rioting in the streets. Jermaine excitedly joins in and starts yelling, breaking windows, and rushing at cops with the group. In the chaos, he knocks down a police officer.

Jermaine knocked down the police officer on purpose.

Jermaine chose to rebel on his own.

### *Purity*

Ajay's coworker tells him about a great group he's a part of called Necrophilia's Secret, a club in which members dedicate their bodies to the group after they die for members to have sex with.

Ajay joins the group to have sex with corpses.

Ajay had sex with the corpses on purpose.

Ajay chose to desecrate the corpses he has sex with.

### *Liberty*

Wilbur is very proud to be American, but his best friend is from Spain, and still a Spanish citizen. Wilbur's dad suggests that it is unpatriotic for the friend not to become an American citizen. Wilbur demands that his friend give up his native citizenship and become an American citizen.

Wilbur takes away his friend's choice on purpose.

Wilbur is responsible for his actions.

### **Intentional (Not responsible):**

### *Harm*

Benny goes out with to a bar one night with a friend, where his friend starts a fight and encourages Benny to join in. Benny excitedly joins in the fight and starts kicking a stranger while he's on the ground.

Benny hurt the stranger on purpose.

Benny chose to hurt the stranger on his own.

### *Fairness*

Floyd works for a politician, who doesn't mind lying or fudging the numbers about his accomplishments. During the election, Floyd distributes slanderous information about the opposing candidate that he believes to be true to help his boss win.

Floyd slandered the other candidate on purpose.

Floyd thought the information was true, so he was not responsible for acting unfairly.

### *Ingroup*

Arnold is part of a secret society. He has a friend who is not part of the society, and who asks him to share some secrets about his group that he isn't supposed to tell. Arnold doesn't want to offend his group, so tells his friend something about the group that he doesn't think is a secret, but which is actually very sensitive information that he wasn't supposed to tell.

Arnold intentionally told his friend information about his group.

Arnold didn't realize the information was secret, so he is not responsible for sharing the secrets.

### *Authority*

Jermaine goes with his friend to a rally protesting the government, where people are shouting and rioting in the streets. Jermaine gets caught up in the mob mentality and starts yelling,

breaking windows, and rushing at cops with the group. In the chaos, he knocks down a police officer.

Jermaine knocked down the police officer on purpose.

Jermaine was affected by the mob mentality and therefore he was not responsible for his actions.

### *Purity*

Ajay's coworker tells him about a great group he's a part of called Necrophilia's Secret, a club in which members dedicate their bodies to the group after they die for members to have sex with.

Ajay joins the group to have sex with corpses.

Ajay had sex with corpses on purpose.

The people consented so Ajay was not responsible for harming them.

### *Liberty*

Wilbur is very proud to be American, but his best friend is from Spain, and still a Spanish citizen. Wilbur's dad suggests that it is unpatriotic for the friend not to become an American citizen. Wilbur demands that his friend give up his native citizenship and become an American citizen.

Wilbur takes away his friend's choice on purpose.

Wilbur doesn't realize how much it means to his friend to be a Spanish citizen, so he is not responsible for violating his friend's choice.

## **Unintentional (Responsible)**

### *Harm*

Benny goes out with to a bar one night with a friend, where his friend starts a fight and encourages Benny to join in. Benny does not join the fight, but accidentally shoves someone to the ground because he wasn't paying attention to where he was going.

Benny accidentally hurt the stranger.

Benny should have been paying attention so he's responsible for harming him.

### *Fairness*

Floyd works for a politician, who doesn't mind lying or fudging the numbers about his accomplishments. During the election, Floyd accidentally distributes untrue, slanderous information about the opposing candidate.

Floyd accidentally damages the reputation of the candidate.

Floyd didn't check his sources, so he is responsible for acting unfairly.

### *Ingroup*

Arnold is part of a secret society. He has a friend who is not part of the society, and who asks him to share some secrets about his group that he isn't supposed to tell. Arnold doesn't want to offend his group, so tells his friend something about the group that he doesn't think is a secret, but which is actually very sensitive information that he wasn't supposed to tell.

Arnold accidentally told his friend information about his group that he didn't realize was secret.

Arnold didn't ask anyone if he could share the information, so he is responsible for sharing the secrets.

### *Authority*

Jermaine goes with his friend to a rally protesting the government, where people are shouting and rioting in the streets. Jermaine does not join in the rioting, but in the chaos he accidentally runs into a police officer and knocks him down.

Jermaine accidentally knocked down the police officer.

Jermaine chose to be at the rally, so he is responsible for his actions.

### *Purity*

Ajay works at a mortuary. His coworker does not respect the dead and is often violent with the corpses. Ajay is not disrespectful, but sometimes he has sex with his girlfriend in the same room as the corpses.

Ajay accidentally did something impure.

Ajay chose to have sex in the same room as the corpses, so he is responsible for his actions.

### *Liberty*

Wilbur is very proud to be American, but his best friend is from Spain, and still a Spanish citizen. Wilbur's dad suggests that it is unpatriotic for the friend not to become an American citizen. Wilbur demands that his friend give up his native citizenship and become an American citizen.

Wilbur thinks he is suggesting the best option, but in doing so, he accidentally takes away his friend's choice.

Wilbur should have asked his friend how he felt, so is responsible for his actions.

### **Unintentional (Not responsible)**

#### *Harm*

Benny goes out with to a bar one night with a friend, where his friend starts a fight and encourages Benny to join in. Benny does not join the fight, but accidentally shoves someone to the ground while he was trying to get away.

Benny accidentally hurt the stranger.

Benny was trying to get away, so he is not responsible for his actions.

### *Fairness*

Floyd works for a politician, who doesn't mind lying or fudging the numbers about his accomplishments. During the election, Floyd accidentally distributes untrue, slanderous information about the opposing candidate.

Floyd accidentally damages the reputation of the candidate.

Floyd thought the information was true, so he was not responsible for acting unfairly.

### *Ingroup*

Arnold is part of a secret society. He has a friend who is not part of the society, and who asks him to share some secrets about his group that he isn't supposed to tell. Arnold doesn't want to offend his group, so tells his friend something about the group that he doesn't think is a secret, but which is actually very sensitive information that he wasn't supposed to tell.

Arnold accidentally told his friend information about his group that he didn't realize was secret.

Arnold didn't know the information was secret, so he is not responsible for his actions.

### *Authority*

Jermaine goes with his friend to a rally protesting the government, where people are shouting and rioting in the streets. Jermaine does not join in the rioting, but in the chaos he was shoved and accidentally runs into a police officer and knocks him down.

Jermaine accidentally knocked down the police officer.

Jermaine was not trying to harm the officer, so he is not responsible for his actions.

### *Purity*

Ajay works at a mortuary. His coworker does not respect the dead and is often violent with the corpses. Ajay is not disrespectful, but sometimes he has sex with his girlfriend in the same room as the corpses.

Ajay accidentally did something impure.

Ajay doesn't think it's gross to have sex near the corpses, so he is not responsible for doing something impure.

### *Liberty*

Wilbur is very proud to be American, but his best friend is from Spain, and still a Spanish citizen. Wilbur's dad suggests that it is unpatriotic for the friend not to become an American citizen. Wilbur demands that his friend give up his native citizenship and become an American citizen.

Wilbur thinks he is suggesting the best option, but in doing so, he accidentally takes away his friend's choice.

Wilbur doesn't realize how much it means to his friend to be a Spanish citizen, so he is not responsible for violating his friend's choice.

### Questions for all vignettes

*Please answer the following true/false questions, based on the information provided above.*

[Focal Character] acted on purpose.

[Focal Character] chose to act the way he did.

*Please answer the following questions based on your interpretation of the scenario above.*

To what extent do *you* believe the following

[from 1 (*not at all*) to 7 (*a great deal*)]

[Focal Character] acted on purpose.

[Focal Character] chose to act the way he did.

Indicate your agreement or disagreement with the following statements

[from 1 (*strongly disagree*) to 7 (*strongly agree*)]

[Focal Character]'s friend is evil

[Focal Character] is evil

If you were the victim of this situation how much would you feel

[from 1 (*not at all*) to 7 (*a great deal*)]

hate toward [Focal Character]?

dislike toward [Focal Character]?

hate toward [Focal Character]'s friend?

dislike toward [Focal Character]'s friend?

Indicate your agreement or disagreement with the following statements

[from 1 (*strongly disagree*) to 7 (*strongly agree*)]

[Focal Character] is infected by his friend's immorality

[Focal Character]'s own immorality is awakened by his friend's

[Focal Character] chooses to act immorally

[Focal Character] is immoral in this situation, but not generally

#### Moral Vitalism Scale

*Please rate your agreement with each of the following items on a scale of 1 (strongly disagree) to 5 (strongly agree).*

1. There are underlying forces of good and evil in this world
2. Either the forces of good or the forces of evil are responsible for most of the events in the world today
3. The forces of good and evil often motivate human behavior
4. People need to be aware of the good and evil that are in this world today
5. Good and evil are aspects of the natural world

#### Evil Measures

*Please rate your agreement with each of the following items on a scale of 1 (strongly disagree) to 5 (strongly agree).*

6. Everyone has evil in them
7. Only some people have evil in them, but it is contagious and might spread to others
8. Only some people have evil in them, and it is not contagious
9. No one has evil in them
10. Evil is a choice
11. People can be evil in the right situation
12. Other (please specify)

#### Moral Foundations Questionnaire

*Part 1. When you decide whether something is right or wrong, to what extent are the following considerations relevant to your thinking? Please rate each statement using this scale*

[0] = not at all relevant (This consideration has nothing to do with my judgments of right and wrong) [1] = not very relevant; [2] = slightly relevant; [3] = somewhat relevant; [4] = very

relevant; [5] = extremely relevant (This is one of the most important factors when I judge right and wrong)

- \_\_\_\_\_ 11. Whether or not someone suffered emotionally
- \_\_\_\_\_ 12. Whether or not some people were treated differently than others
- \_\_\_\_\_ 13. Whether or not someone's action showed love for his or her country
- \_\_\_\_\_ 14. Whether or not someone showed a lack of respect for authority
- \_\_\_\_\_ 15. Whether or not someone violated standards of purity and decency
- \_\_\_\_\_ 16. Whether or not someone was good at math
- \_\_\_\_\_ 17. Whether or not someone cared for someone weak or vulnerable
- \_\_\_\_\_ 18. Whether or not someone acted unfairly
- \_\_\_\_\_ 19. Whether or not someone did something to betray his or her group
- \_\_\_\_\_ 20. Whether or not someone conformed to the traditions of society
- \_\_\_\_\_ 21. Whether or not someone did something disgusting
- \_\_\_\_\_ 22. Whether or not someone was cruel
- \_\_\_\_\_ 23. Whether or not someone was denied his or her rights
- \_\_\_\_\_ 24. Whether or not someone showed a lack of loyalty
- \_\_\_\_\_ 25. Whether or not an action caused chaos or disorder
- \_\_\_\_\_ 26. Whether or not someone acted in a way that God would approve of

*Part 2. Please read the following sentences and indicate your agreement or disagreement:*

[0]	[1]	[2]	[3]	[4]	[5]
Strongly disagree	Moderately disagree	Slightly disagree	Slightly agree	Moderately agree	Strongly agree

- \_\_\_\_\_27. Compassion for those who are suffering is the most crucial virtue.
- \_\_\_\_\_28. When the government makes laws, the number one principle should be ensuring that everyone is treated fairly.
- \_\_\_\_\_29. I am proud of my country's history.
- \_\_\_\_\_30. Respect for authority is something all children need to learn.
- \_\_\_\_\_31. People should not do things that are disgusting, even if no one is harmed.
- \_\_\_\_\_32. It is better to do good than to do bad.
- \_\_\_\_\_33. One of the worst things a person could do is hurt a defenseless animal.
- \_\_\_\_\_34. Justice is the most important requirement for a society.
- \_\_\_\_\_35. People should be loyal to their family members, even when they have done something wrong.
- \_\_\_\_\_36. Men and women each have different roles to play in society.
- \_\_\_\_\_37. I would call some acts wrong on the grounds that they are unnatural.
- \_\_\_\_\_38. It can never be right to kill a human being.
- \_\_\_\_\_39. I think it's morally wrong that rich children inherit a lot of money while poor children inherit nothing.
- \_\_\_\_\_40. It is more important to be a team player than to express oneself.
- \_\_\_\_\_41. If I were a soldier and disagreed with my commanding officer's orders, I would obey anyway because that is my duty.
- \_\_\_\_\_42. Chastity is an important and valuable virtue.

#### Bad People

*Please read the following statements and indicate your agreement or disagreement:*

[0]            [1]            [2]            [3]            [4]            [5]

Strongly	Moderately	Slightly	Slightly	Moderately	Strongly
disagree	disagree	disagree	agree	agree	agree

Bad people should be...

1. punished
2. harmed so they can't harm others
3. harmed so they can't harm the self
4. left alone
5. avoided so they can't harm others
6. avoided so they can't harm the self

## Demographics

Demographic Information

Gender: Male Female Other

Age (in years):

Ethnicity (choose all that apply):

White/Caucasian

Black/African American

Asian/Asian American

Hispanic/Latinx

Native American

Pacific Islander

Other

## Appendix C

Study 2 – Two prompts are presented below: one that was presented to answer questions about the self, and the other to answer questions about another person. The questions were tailored to condition (e.g. “how intensely did you feel hate toward the person?”) but to save space only one set of questions is presented here. The remaining questions are available upon request.

### Self Conditions

*Think of a time you experienced hate[dislike] toward a group[individual].*

What[who] was the group[individual]?

Describe what it was that made you experience hate toward the group at the time.

### Other Conditions

*Think of a time someone you know experienced hate[dislike] toward a group[individual].*

What is name of the person who experienced hate[dislike]?

What[who] was the group[individual]?

Describe what it was that made you experience hate toward the group at the time.

Questions for all participants

How intensely did you[other person] feel...?

[from 1 (*not at all*) to 7 (*a great deal*)]

hate toward the group[individual]

dislike toward the group[individual]

threatened by the group[individual]

the group[individual] was evil

the group[individual] did something evil

the group[individual] was immoral

the group[individual] did something on purpose

the group[individual] was purposely bad

the group[individual] is responsible for their actions

the group[individual] is responsible for the way they are

Please rate your agreement with each of the following items on a scale of 1 (strongly disagree) to 7 (strongly agree).

By experiencing hate[dislike] toward the group[individual], I[other] activated evil in myself

By experiencing hate[dislike] toward the group[individual], I[other] could have infected others with evil

By experiencing hate[dislike] toward the group[individual], I[other] could have harmed others

I[other] resisted feeling hate[dislike] toward the group[individual]

It was wrong to experience hate[dislike] toward this group[individual]

It is wrong to feel hate[dislike] toward a group[individual]

After you experienced dislike toward the group, to what extent did you want to...

from 1 (*not at all*) to 7 (*a great deal*)

confront or attack this group[individual] because you wanted to cause them pain or harm

confront or attack this group[individual] because you wanted to protect yourself from further harm

confront or attack this group[individual] to avoid the consequences of staying away from them

stay away from this group[individual] because you wanted your absence to cause them

pain or harm

stay away from this group[individual] because you wanted to protect yourself from further harm

stay away from this group[individual] to avoid the consequences of confronting or attacking them

Please indicate on a scale of 0 to 100 how much you wanted to...

Approach the group[individual]

Avoid the group[individual]

[See Appendix B for the remaining scales]

Moral Vitalism Scale

Evil Measures

Moral Foundations Questionnaire

Bad People Measures

Demographics

## Appendix D



---

**To:** Denise R Beike  
MEMH 202D

**From:** Douglas James Adams, Chair  
IRB Committee

**Date:** 03/12/2018

**Action:** **Exemption Granted**

**Action Date:** 03/12/2018

**Protocol #:** 1802103935

**Study Title:** Impressions of Others' Actions

The above-referenced protocol has been determined to be exempt.

If you wish to make any modifications in the approved protocol that may affect the level of risk to your participants, you must seek approval prior to implementing those changes. All modifications must provide sufficient detail to assess the impact of the change.

If you have any questions or need any assistance from the IRB, please contact the IRB Coordinator at 109 MLKG Building, 5-2208, or [irb@uark.edu](mailto:irb@uark.edu).

cc: Carmen Merrick Merrick, Investigator