

Online Hate and its Routes to Aggression: A Research Agenda

Argha Ray
Iowa State University
argharay@iastate.edu

Joey F. George
Iowa State University
jfgeorge@iastate.edu

Abstract

The Internet has provided human civilization hitherto unimaginable tools with which to connect, communicate and coexist. Unfortunately, those same tools are being maneuvered to spread ethno-racial and religious hatred. The internet is now replete with chat rooms, web pages, discussion boards, forums, videos, music, and games that actively promote violence against outgroups. To that extent, it is almost impossible now to navigate through the internet without encountering hateful ideologies and propaganda that deepen societal fissures and instigate violence. Though journalism has put the spotlight on the link between online radicalization and real-world hate crimes, empirical evidence on these claims is largely lacking. The existing evidence is merely anecdotal cutting across multiple scientific disciplines. This paper lays out an operationalizable research design that may shed more light on this causal link between online hate and hate crimes. We review the existing literature on online hate and radicalization and propose that the General Aggression Model may be a model of choice to empirically investigate the link between online hate and offline violence.

1. Introduction

Just prior to opening fire on March 15th, 2019 outside a mosque in Christchurch, New Zealand, Brenton Tarrant published an 87 page long manifesto on Twitter and 8chan justifying the attack and his anti-immigrant and anti-Muslim sentiments. He also live-streamed the attack on Facebook while killing 51 people and injuring 49 [25]. His act also inspired Patrick Crusius, who entered a Walmart Supercenter in El Paso, Texas on August 3rd, 2019 carrying a semi-automatic rifle and killed 23, while injuring another 23 before surrendering to the police [25]. This phenomenon is not new but ongoing for some time now. On January 29, 2017 Alexandre Bissonnette entered a mosque in Quebec City, Canada, and fired indiscriminately on gathered worshippers, killing 6

and injuring 19. He followed several profiles from French extremist groups on his Facebook page [24]. A few years ago, on July 22, 2011, Anders Breivik killed 77 people and injured 319 in Norway in what was the deadliest act of terrorism on Scandinavian soil. The internet intensified his violent radicalization and helped him to acquire skills in making a fertilizer bomb [69]. But an incident that perhaps shook the collective American conscience, was the bombing of the Alfred P. Murrah Federal Building in Oklahoma City, Oklahoma, on April 19, 1995 killing 168 people. Timothy McVeigh used a truck bomb strikingly similar to one described in a dystopian novel, *The Turner Diaries*, written by William Pierce and published some 40 years ago. Although, Timothy McVeigh himself was never radicalized online, the book has also inspired recent terror manifestos and its online versions are actively promoted and regularly shared as required text on online hate forums to motivate copycat acts.

The above examples have two interlinked threads, online hate and perpetration of offline violence. This link between online hate and random acts of unrestrained aggression with an intent to kill is only anecdotal and speculatively theorized, mainly in journalistic commentaries. But it has intuitive appeal. One of the earliest researchers on online hate, British sociologist Les Back, eerily predicted some 20 years back “The real danger is perhaps that in the ‘informational age’ isolated acts of racist terrorism may become commonplace” [5] (p. 642), while examining digital technology usage within extremist movements. Hence, our research question is as follows: Does exposure to hate online lead to harmful aggression? The existing literature on online hate is limited on this causal question. Therefore, we look at the mature literature on media violence from social psychology to operationalize an experimental paradigm to investigate the causal link between online hate and aggression. Specifically, we use the *General Aggression Model* (GAM) to inform our exploratory

quest on a research question that has far reaching societal impact. The paper is organized as follows. First, we briefly review the literatures on online hate and radicalization. We then provide a theoretical background on GAM, emphasizing its relevance to understand the phenomena at hand.

2. Literature Review

Online hate is a cross disciplinary subject of research. Related concepts include cyber racism, cyber hate, cyber bigotry, online racism, micro aggression etc. Thus far the focus of research questions has been on how individuals use the internet to disseminate racist messages to validate and endorse their views, and how extremist groups widen their support base and strengthen commitment of existing members [41]. The scientific disciplines include sociology, anthropology, political science, criminology, social psychology, communication studies, cultural studies, and information science. The scientific methodologies include content analysis, thematic analysis, critical discourse analysis, virtual ethnography, grounded theory, online surveys, web mining, computerized linguistic analysis and social network analysis [41]. The literature is predominantly qualitative with no experimental literature. A recent literature review by Bliuc et al. [13] opines that further research in this area should “utilise a wider array of methods than qualitative text analysis,” and “investigate how online racists radicalise isolated individuals” (p. 85). We take a social psychological viewpoint to address the call for research situated within an American context.

2.1 Online Hate

Hate went online in America sometimes in early March 1984, when George Dietz launched *Info. International Network* or *Liberty Bell Net* on a public bulletin board system (BBS). Next came Louis Beam’s *Aryan Liberty Net* sometime in spring of 1984, followed by Tom Metzger’s *White Aryan Resistance BBS* launched sometimes in late 1984 or early 1985 in Fallbrook, California. With time BBS’s continued to rise but either disappeared or moved to the USENET. Finally, with the advent of the World Wide Web, Don Black set up *Stromfront* in May 1995 [11]. Plagued by criminal prosecutions and civil lawsuits under tort laws for connections with violence, the internet became the cornerstone of the “leaderless resistance” strategy for extremist groups because of its low cost of distribution, ability to tailor message to specific

audience, and the ease of global reach [44]. The number of hate websites have since kept rising.

Unlike other liberal democracies, such hate speech on the internet enjoys near absolute protection under the 1st Amendment. The Supreme Court identifies certain categories as unprotected speech viz. obscenity, defamation, fraud, incitement, fighting words, true threats, speech integral to criminal conduct, and child pornography. Some of these categories which are applicable to hate speech on the internet have been struck down by the court in successive rulings. In *Cohen v. California* (1971) the court observed that an individual could not be criminally prosecuted for merely wearing a jacket bearing profanity in reference to conscription in a corridor of a Los Angeles courthouse, overturning the fighting words doctrine expressed in *Chaplinsky v. New Hampshire* (1942). Similarly, the court took a narrow view on fighting words in *R.A.V. v. City of St. Paul* (1992), where it disregarded St. Paul’s *Bias-Motivated Crime Ordinance* as sound law after a teenager burned a cross in the lawn of an African American family. On the incitement to imminent lawless action doctrine, the court overturned the conviction of Clarence Brandenburg in *Brandenburg v. Ohio* (1969), a case where several armed men in robes and hoods set a wooden cross ablaze while uttering phrases that targeted African Americans and Jews. Finally, the court set aside the conviction of anti-war protester Robert Watts in *Watts v. United States* (1969) on grounds that his utterance, “If they ever make me carry a rifle the first man I want to get in my sights is L. B. J.,” was political hyperbole and did not constitute a true threat to the life of President Lyndon B. Johnson. In 1997, the court also granted broad 1st Amendment protection to internet-based communication while contending that “the content of the Internet is as diverse as human thought” in *Reno v. American Civil Liberties Union* (1997). Though a signatory to the *International Convention on the Elimination of All Forms of Racial Discrimination*, 1965, the United States expressed reservation and did not accept obligation under Article 4 of the convention to “condemn all propaganda and all organizations which are based on ideas or theories of superiority of one race or group of persons of one colour or ethnic origin, or which attempt to justify or promote racial hatred and discrimination in any form” [67].

The near absolute guarantee has created a “safe haven for foreign haters” [3]. Ernst Zündel, a German citizen residing in Canada, faced criminal liability for

Theories	Description
Staircase to Terrorism	Although, a vast majority of people feel deprived and unfairly treated, most people remain on the ground floor, and some individuals climb up the metaphorical staircase and commit acts of terrorism.
Pathway Model	Individuals take different pathways to terrorism influenced by various factors. Four stages include "It's not right," "It's not fair," "It's your fault," and "You're evil."
Theory of Joining Extremist Groups	There are four processes that lead a person to join an Islamic extremist group viz. cognitive opening, religious seeking, frame alignment, and socialization.
NYPD Radicalization Process	Four stages of radicalization include pre-radicalization, self-identification, indoctrination, and jihadization.
Four Prongs Model	Radicalization emerges from the interplay of four factors viz. sense of moral outrage, single narrative, personal experience, and mobilization through networks.
Integrated Threat Theory	Prejudice and negative attitudes towards outgroups are explained by four types of threats viz. realistic threat, symbolic threat, negative stereotype, and intergroup anxiety.
Compensatory Control Theory	People compensate for a perceived loss of personal control by endorsing external systems such as religion, government, or superstition.
Goal Systems Theory	Extremism is an expression of goal commitment under uncertainty. Zeal is a direct expression of goal commitment, whereas deviant behavior occurs under high commitment because of the perceived instrumentality of such behavior towards the goal.
Uncertainty-Identity Theory	People reduce uncertainty by identifying with radical group with strong restrictions on acceptable attitudes and behavior.
Mindset and Worldview	Mindset combined with a specific worldview create a climate for violent extremism. Mindset includes vulnerabilities and propensities, while worldview is situational.
Reactive Approach Motivation Theory	Personal uncertainty causes anxiety, and that anxiety draws people towards extremes because such extremes activate approach-motivated states that automatically reduce anxiety.
The Devoted Actor Model	Devoted actors are deontic agents who are devoid of instrumental rationality and are willing to sacrifice for an ingroup with whom their personal sense of identity is fused.
The Two-Pyramids Model	Radicalization of opinion is separate from radicalization of action.
Quest for Significance or "3N"	Needs i.e. individual motivation, narratives i.e. ideological justification of violence, and networks i.e. group processes lead to radicalization.

Table 1: Competing Theories of Radicalization

his holocaust denial website under Section 131 of the *German Criminal Code* in Germany, and also faced similar charges in Canada under Section 13 of the *Canadian Human Rights Act*. Shortly, thereafter he migrated his website to California [18]. The website promoted Holocaust denial. Similar websites and social media channels such as 8chan, Gab, Telegram etc. have sprung up from time to time to disseminate “politically incorrect” messaging. When some social media firms or web infrastructure providers de-platform such explicit content, the online communities of hate relocate to other more welcoming platforms. The content includes manifestos of mass shooters, livestreams of mass shooting, do it yourself weapons design, and hit lists of prominent politicians, journalists, celebrities, and members of minority communities. The online subculture of hate uses slick aesthetics to make violence look more appealing and encourage its transnational userbase to follow the footsteps of mass shooters, who are deified as saints. This widespread proliferation of violent content on the internet has lowered the barrier to entry into the extremist ecosystem and has complicated the efforts of law enforcement agencies to predict when someone

will evolve from wishing others dead to actually perpetrating the act [48]. But can mere promotion of violence online lead to radicalization and mass violence?

2.2 Radicalization

The psychological literature on radicalization, or specifically Islamic radicalization, gained traction in academic circles only after the 9\11 terrorist attack and is variedly conceptualized as “what goes on before the bomb goes off” [52] (p. 4). Some of the competing psychological theories that have been proposed to understand radicalization are: *The Staircase to Terrorism* [50], *Pathway Model* [15], *Theory of Joining Extremist Groups* [73], *NYPD Radicalization Process* [58], *Four Prongs Model* [56], *Integrated Threat Theory* [60], *Compensatory Control Theory* [42], *Goal Systems Theory* [43], *Uncertainty-Identity Theory* [39], *Mindset and Worldview* [14], *Reactive Approach Motivation Theory* [47], *The Devoted Actor Model* [4], *The Two-Pyramids Model* [46], and *Quest for Significance* or “3N” [72]. Table 1 provides a brief description of the theories. None of the above approaches fully explain radicalization and are difficult to empirically operationalize to tease out

Theories	Description
Social Learning Theory	People acquire aggressive responses by direct experience or by observing others.
Script Theory	Children exposed to violence in mass media learn aggressive scripts that guide future behavior.
Cognitive Neoassociation Theory	Aversive events produce negative affect which stimulates various thoughts, memories, and physiological responses associated with aggressive behavior.
Excitation Transfer Theory	If two arousing events happen in quick succession, arousal from the earlier event may be misattributed to the subsequent event.
Social Interaction Theory	Coercive actions can produce change in the target's behavior.
Social Information Processing Theory	Aggressive children attend, perceive, interpret, and make decisions about information in ways that increase their likelihood to aggress.

Table 2: Domain Specific Theories of Aggression

causality. Gøtzsche-Astrup [37] recently reviewed the existing literature on radicalization and identified some central overlapping ideas in proposed theories of radicalization that have some empirical support. He noted that negative life experiences lead to fundamental uncertainty or loss of significance, which spur a search for identity by shifting towards groups with strong norms and ideals, including sacred values that enable acts of terrorism. He also lamented that there are minimal experimental investigations to support causal claims. Borum explains “The reasons for this relative lack of empirical inquiry are varied, but include difficulty gaining access to terrorists as subjects for research (because they may be dead, underground, or incarcerated) and inability of many academic researchers to access classified data or information” [16] (p. 66). Considering the current state of the literature our goals are more modest for our current research agenda. Instead of understanding the process of radicalization, we focus on whether online hate leads to reactive laboratory aggression. We now take recourse to the mature media violence literature to inform our pursuit.

3. Conceptual Model

The *General Aggression Model* (GAM) is a meta-theory that is a theoretical integration of several domain specific theories of aggression. Though it came out of the media violence literature, specifically the impact of violent video games, it is a model that accounts for aggression in general [2]. It encapsulates *Social Learning Theory* [6], *Script Theory* [40], *Cognitive Neoassociation Theory* [9], *Excitation Transfer Theory* [74], *Social Interaction Theory* [65], *Social Information Processing Theory* [31], and also the impact of systematic desensitization [23]. Table 2 provides a brief description of the theories. The above theories make up the different pieces of GAM. The newest version of GAM released in 2018 [2] accounts for both distal and proximal causal factors for aggression. Figure 1 reproduces the same. The distal factors are concerned with the changes in personality traits as a result of repeated exposure to violence. The

proximal factors of the model have inputs, three routes, and outcomes.

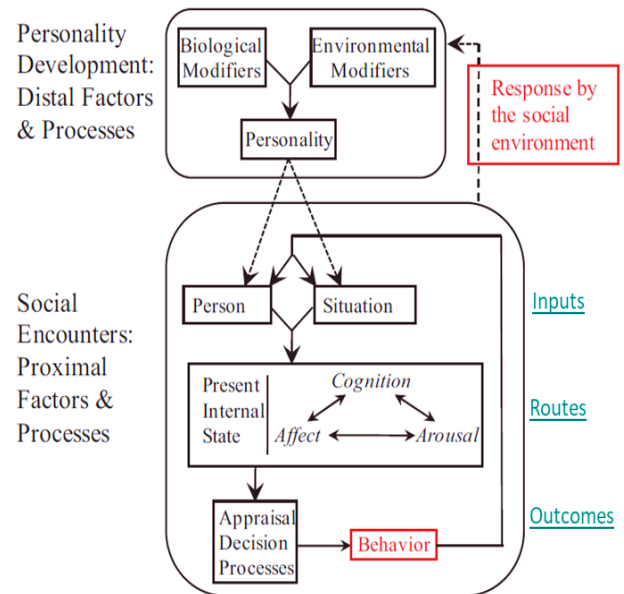


Figure 1: Adapted General Aggression Model [2]

We posit that if an ordinary person with strong sense of ethnic, racial, religious, or national identity is exposed to stimuli that evokes a sense of victimization for the ingroup and promotes violence towards the outgroup, the person in the situation will act aggressively. These input variables will influence the final behavioral outcome through three internal states that occur simultaneously. The persons will be cognitively predisposed to pay more attention to the violent cues in the stimuli. Concurrently, the person will be in a state of excitatory arousal and will experience negative affect towards the outgroup. The internal states will cloud judgment of the person and will lead to quick automatic appraisal of the stimuli as threatening towards the ingroup. This impulsive decision making will cause the person to behave aggressively towards the outgroup. Reactions

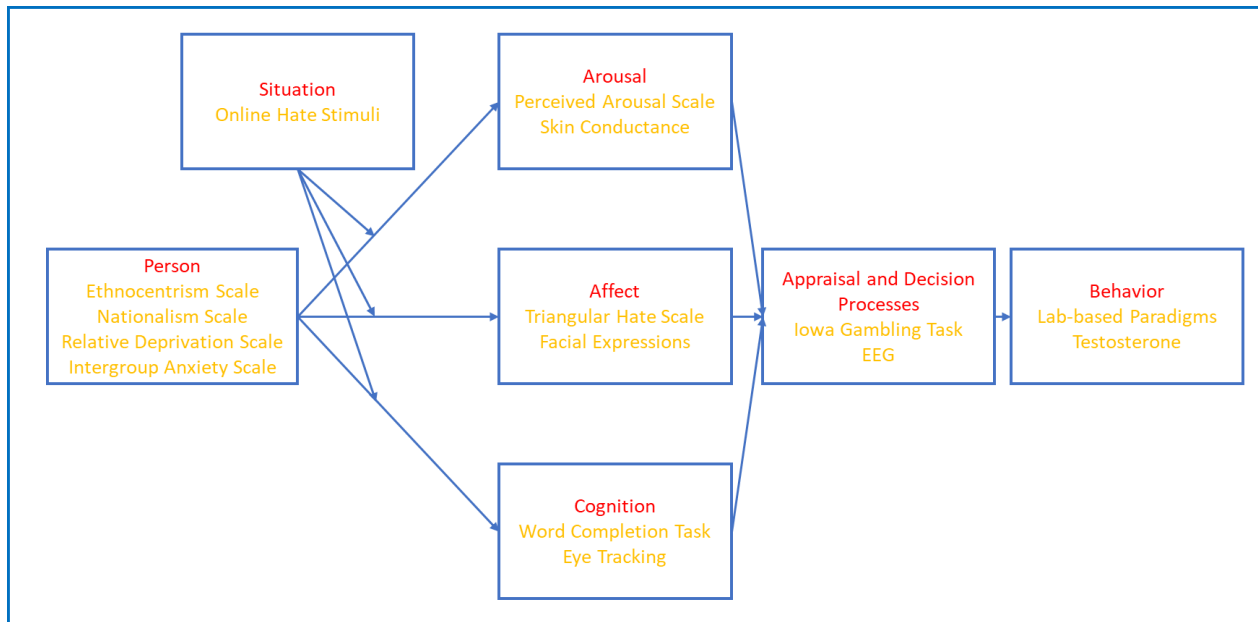


Figure 2: Testable model of GAM

and counter reactions lead to a downward spiral in the aggression cycle. Being repeatedly caught in the downward spiral may accentuate chronic aggressive behavior by changing traits and how the person perceives a new situation. Such distal processes may explain the process of radicalization as it happens over a period of time. For our present purposes, we only focus on how the proximal factors of person, situation, cognition, arousal, affect, and appraisal may elicit aggressive action under laboratory settings. If there is evidence from such a cross sectional study linking online hate to aggression, daily diary studies could be a next step to understand the process of radicalization itself. We propose a testable model of the proximal factors of GAM along with operationalizations in Figure 2 that can be applied to the context of online hate. We now delve deeper into each component of GAM and also indicate how each component of GAM can be operationalized within the laboratory using either psychometric or NeuroIS techniques.

3.1 Person

Person factors include all stable characteristics such as personality traits, attitudes, and genetic predispositions that the person brings to the situation. Research on radicalization has still not been able to accurately profile a “terrorist personality.” The point of view among members of the general public and earlier psychologists that terrorists suffer from psychopathology and paranoia has been discredited. Instead, they are ordinary and unremarkable. The

involvement in terrorism is a gradual process. In a review on the psychological processes of Jihadi radicalization, Silke [59] notes that education, economic background, and marital status has no clear link with membership in extremists’ organizations. But terrorists are predominantly male and young teenagers or in their early twenties. They have an acute sense of perceived deprivation and injustice, value the status of being “freedom fighters,” and have appropriate social networks. Another vital factor is Social Identity, and in case of Jihadi terrorists, it is their Islamic and Terror Group (e.g. Hamas, al-Qaeda) Identity more than national or ethnic identity. Although, the mechanisms of Jihadi radicalization may well be different than domestic radicalization the psychological processes leading to a strong feeling of anxiety or uncertainty, a strong sense of deprivation, and a strong sense of social identity as noted in the radicalization literature may well be the same. These psychological processes may be measured using the Ethnocentrism, Nationalism, Relative Deprivation, and Intergroup Anxiety scales.

3.2 Situation

Situational factors such as aggressive cues, provocation, and frustration leads to aggression. Aggressive cues such as weapons or violent media prime aggressive scripts in memory. Provocations such as insults or slights are a key cause of aggression as also is perceived injustice. Frustrations are an impediment to goal attainment and lead to either

aggression on the agent that caused the frustration or displaced aggression. Online hate is violent towards the outgroup, provoke perceptions of injustice and discrimination of the ingroup, and frustrate by advancing notions of lost privilege and systematic replacement. Stimuli that calls for violence towards groups responsible for perceived deprivation of the ingroup may lead to reactive aggression as it may lead to excitatory arousal, evoke negative affect towards the outgroup, and activate aggressive cognitive schemata in memory by selective attention to violent cues.

3.3 Arousal

Physiological arousal facilitates and intensifies aggression in the presence of aggressive cues. Residual excitation from an irrelevant source can intensify aggression on another target through a process of excitation transfer as it impairs cognitive processes needed to disinhibit aggression [74]. Electrodermal activity is an indicator of arousal. Vigouroux and Féré demonstrated the effect of psychological variables on the exosomatic measure of human electrodermal activity as early as 1880s while working in the laboratory of French neurologist Jean Charcot. The skin is a protective barrier that aids thermoregulation through the production of sweat by the eccrine sweat glands. Those located in the palmar and plantar surfaces are more responsive to psychological than thermal stimuli. Sweat is secreted in the glands depending on the degree of activation of the sympathetic nervous system and cause changes in skin conductance levels [28]. Arousal can be measured using the Perceived Arousal scale or changes in Skin Conductance.

3.4 Affect

Intergroup emotions have been theorized to aid in the process of radicalization and political violence [68]. Commentary on emotions traces back to antiquity when Aristotle distinguished among fifteen emotional states in his book *Rhetoric*. Gendron and Barrett [36] identifies three traditions in the hugely debated emotion literature: the basic emotions perspective, the appraisal perspective, and the psychological constructionist perspective. The basic emotions perspective can be directly attributed to Charles Darwin's 1872 book *The Expression of the Emotions in Man and Animals*. Over the years the list of basic emotions has been shortened, and Ekman identifies them as contempt, anger, disgust, fear,

happiness, sadness, and surprise [32]. Theorists have suggested hate to be a combination of emotions. Sternberg [61] proposed that hate is a substantial contributing factor to terrorism, massacres, and genocide. His *Duplex Theory of Hate* theorized a triarchic model of hatred, composed of emotions that include anger or fear, contempt, and disgust forming the triangle of hate. The combination of these three components can lead to seven types of hatred: cold, cool, hot, simmering, boiling, seething, and burning hatred. Emotions are communicated using facial expressions in both human and non-human primates and emerge from the coordinated activity of the amygdala and multiple interconnected cortical and sub cortical motor areas. Hate can be measured using the Triangular Hate scale or Facial Expressions.

3.5 Cognition

Aggression increases due to chronic accessibility of aggressive scripts acquired through social learning. Scripts are well rehearsed concepts in semantic memory linked together in associative networks. Exposure to stimulus activates scripts to interpret social information and guide action tendencies [40]. *Social Information Processing Theory* [31] of aggression in developmental psychology notes that the first step in social cognition is encoding of situational cues by selectively attending to hostile cues and ignoring other cues leading to aggressive social response in ambiguous situations. Although our visual system carries out an exhaustive extraction of visual information from the environment, our behavior is guided by only a small subset of that information which is in our attention span. Moore and Zirnsak [51] states that this "aspect of visually guided behavior is referred to as selective visual attention, and it is among the more fundamental cognitive functions" (p. 48). Eye movements reflect engagement of attention during encoding of visual information as the first step in social cognition. Eye movements had been remarked upon for millennia, with Roman philosopher Cicero describing the eye as 'window to the soul,' but experimental investigation was initiated by Scottish American physician William Charles Wells only in 1792 [70]. The first evidence that eye movements might facilitate attention came from a psychophysical study by Crovitz and Daves in 1962 [27]. Attention stems when a frontal-parietal control network interacts in rhythmic synchrony with the visual cortex. Accessibility of aggressive scripts can be measured

Paradigms	Description
Competitive Reaction Time Task	Selecting the intensity of sound blasts to an individual.
Bobo-doll Modeling Task	Aggressive behavior of children after watching physical abuse of a bobo doll.
Teacher/Learner Task	Intensity and duration of electrical shocks delivered to an individual, who is a learner, for an incorrect response.
Essay Evaluation	Number of electrical shocks delivered to an evaluator, who negatively evaluated the written solution to a problem solving task, by the participant.
Point-Subtraction Aggression Paradigm	Subtracting money or blasting white noise to an individual in response to provocation.
Cold Pressor Task	Choosing how long an individual have to hold their hand in ice water.
Tangram Help/Hurt Task	Selecting difficult Tangram puzzles, which decreases the probability of an individual in obtaining a desired reward.
Hot Sauce Paradigm	Choosing how much hot sauce an individual will have to consume.
Uncomfortable Pose Task	Choosing how long an individual has to hold a physically uncomfortable body posture.
Voodoo Doll Task	Choosing how many pins be pierced into a representation of an individual to symbolically inflict harm.
Fight or Escape Paradigm	Choosing the volume of a sound blast to an individual in response to provocation.
Negative Evaluation Task	Evaluating an individual, which deceases their likelihood to obtain a desired goal.

Table 3: Lab Based Aggression Paradigms

using the Word Completion Task and attention by using Eye Tracking.

3.6 Appraisal and Decision Processes

Encounter with a situation leads to an immediate initial appraisal. This initial appraisal may lead to an automatic impulsive action or a reappraisal of the situation. During reappraisal alternative explanations of the situation are considered, and that leads to thoughtful action. Strack and Deutsch's dual process *Reflective Impulsive Model* [62] states that social behavior is controlled by two interacting systems. While the reflective system generates behavioral decisions that are based on facts, values and noetic weighing of probable consequences, the impulsive system elicits behavior based on spreading activation of associative links and motivational orientations. Two distinct motivational orientations that guide processing of information and activation of behavior are approach and avoidance. Approach orientation is to fight the target and avoidance orientation is flight away from the target. Previous research on approach and avoidance records that the amygdala is responsible for the determination of motivational relevance, and areas in the prefrontal cortex play a role in amygdala activation when we process motivational information. Frontal cortical asymmetry in approach and avoidance motivations was suggested as early as 1939 by German neurologist and psychiatrist Kurt Goldstein. The left dorsolateral cortex is involved in approach processes and the right dorsolateral prefrontal cortex is involved in withdrawal processes [38]. Impulsive decision making can be measured using the Iowa

Gambling task and approach motivation can be measured using EEG.

3.7 Behavior

Aggression is defined as behavior that is carried out with an intent to cause harm to the target, who is motivated to avoid the harm [8]. For ethical, legal, and practical issues involved in inciting aggression within laboratory settings, social scientists have developed several lab-based aggression paradigms. Notable lab-based paradigms include *Taylor Aggression Paradigm* or *Competitive Reaction Time Task* [64, 33], *Bobo-doll Modeling Task* [7], *Teacher/Learner Task* [19], *Essay Evaluation* [10], *Point-Subtraction Aggression Paradigm* [22], *Cold Pressor Task* [53], *Tangram Help/Hurt Task* [57], *Hot Sauce Paradigm* [45], *Uncomfortable Pose Task* [34], *Voodoo Doll Task* [29], *Fight or Escape Paradigm* [12], and *Negative Evaluation Task* [30]. Table 3 provides a brief description of the paradigms. Additionally, hormones and social behavior has a bidirectional relationship. In 1849, German physiologist and zoologist Arnold Berthold performed the first known experiment in behavioral neuroendocrinology. He removed the testes of male roosters and found castration led to lesser mating and aggression. He then implanted the testes in the body cavity and the male phenotype redeveloped mating and aggressive behavior. This led him to conclude that the testes must secrete some substance in the bloodstream which influences the typical behavior of the male phenotype. That substance is the steroid hormone testosterone, synthesized by the Leydig cells of the testes and to a lesser extent by the adrenal cortex. Change in testosterone concentrations

positively predict ongoing or future human aggression [20]. Behavior may be measured using either the lab-based aggression paradigms or by changes in Testosterone levels.

3. General Discussion

The link between online hate and hate crimes is unexplored. Considering the current state of the literature crisscrossing multiple disciplines, we fall back on a separate tried and tested literature on GAM that may help unravel causality of the phenomenon. The development of GAM started with what was originally known as the *General Affective Aggression Model* in 1995 and was later renamed to *General Aggression Model* in 2002 [2]. It gained prominence as a theoretical model to explain the link between violent video games and real life violence in the aftermath of the Columbine school shooting where two teenagers enacted the gameplay of the video game Doom in real life. The shootings started an unresolved public debate about video game violence and triggered US Senate Committee hearings where several researchers testified about the harmful effects of exposing youths to violent video games. Since then GAM has been refined with the latest model released in 2018. However, GAM has never been applied in the context of online hate even though Anderson and Bushman [2] recently speculated that “there is considerable need for additional research on how violent screen media might affect aggression against outgroups (especially but not exclusively minorities) mediated by changes in stereotypes and prejudices. News media and hate speech in particular seem types of violent screen media in need of additional effects research” [2] (p. 407). Although a meta-theory that is a theoretical integration of several domain specific theories of aggression, GAM in its present form falls short on the question of online hate. Hence, we supplement GAM with even more domain specific theories in three areas. We bring in the theories of radicalization for the Person component of GAM, the *Duplex Theory of Hate* for the Affect component, and the *Reflective Impulsive Model* for the Appraisal and Decision Processes component. We retain the rest of the GAM components as is with the Situation component of GAM ideated as exposure to online hate instead of violent video games. The paper also portrays how the various components of GAM can be operationalized and measured using either psychometric or NeuroIS techniques. It will need a series of laboratory studies, both cross sectional and daily diary studies, to achieve

complete operationalization of GAM applied in the context of online hate and radicalization with both psychometric and neuroscience techniques. For immediate operationalization we envisage an initial cross section study where we want to test the reaction of individuals when exposed to online hate and if that reaction is aggressive in nature. Though our current agenda is more modest in terms of priming laboratory aggression on account of exposure to online hate, GAM holds promise to also untangle the distal causes of aggression i.e. the process of radicalization itself. Anderson and Bushman [2] describe that repeated exposure to violent media alters both psychological and physiological make up the same way as smoking a single cigarette does not cause lung cancer but smoking a pack daily over the years leads to cancer. If the causal link between online hate and hate crimes can be established within the laboratory, it may have wide ranging public policy implications. In particular, it may call into question the near absolute protections granted in the 1st Amendment to hate speech online and offline, and whether laws in other liberal democracies censoring such speech really amounts to infringement on free speech?

4. References

- [1] Anderson, C. A., & Bushman, B. J.. Human aggression. *Annual Review of Psychology*, 53(1), 2002, pp. 27–51.
- [2] Anderson, C. A., & Bushman, B. J. Media violence and the general aggression model. *Journal of Social Issues*, 74(2), 2018, pp. 386–413.
- [3] Anti-Defamation League. *Combating extremism in cyberspace: The legal issues affecting internet hate speech*. New York: ADL, 2000.
- [4] Atran, S. The devoted actor: Unconditional commitment and intractable conflict across cultures. *Current Anthropology*, 57(S13), 2016, pp. 192–203.
- [5] Back, L. Aryans Reading Adorno: Cyber-Culture and Twenty-First- Century Racism. *Ethnic and Racial Studies*, 25(4), 2002, pp. 628–651.
- [6] Bandura A. *Social Learning Theory*. New York: Prentice Hall, 1977.
- [7] Bandura, A. *Aggression: A social learning analysis*. Oxford, England: Prentice-Hall, 1973.
- [8] Baron, R. A. and Richardson, D. R. *Human aggression*. 2nd ed. New York: Plenum Press, 1994.

- [9] Berkowitz L. On the formation and regulation of anger and aggression: a cognitive neoassociationistic analysis. *Am. Psychol.* 45, 1990, pp. 494–503.
- [10] Berkowitz, L. *Aggression: A social psychological analysis*. New York: McGraw-Hill, 1962.
- [11] Berlet, C. *When Hate Went Online.* Northeast Sociological Association, Spring Conference. Fairfield, CT: Sacred Heart University, 2010.
- [12] Beyer, F., Buades-Rotger, M., Claes, M., Krämer, U M. Hit or Run: Exploring Aggressive and Avoidant Reactions to Interpersonal Provocation Using a Novel Fight-or-Escape Paradigm (FOE). *Frontiers in Behavioral Neuroscience*, 11, 2017.
- [13] Bliuc, A.-M., Faulkner, N., Jakubowicz, A., & McGarty, C. Online networks of racial hate: A systematic review of 10 years of research on cyber-racism. *Computers in Human Behavior*, 87, 2018, pp. 75–86.
- [14] Borum, R. Psychological vulnerabilities and propensities for involvement in violent extremism. *Behavioral Sciences & the Law*, 32(3), 2014, pp. 286–305.
- [15] Borum, R. *Understanding the Terrorist Mindset*, FBI Law Enforcement Bulletin, 2003, pp. 7–10.
- [16] Borum, R. *Psychology of terrorism*. Tampa: University of South Florida, 2004.
- [17] *Brandenburg v. Ohio*, 395 U.S. 444 (1969)
- [18] Breckheimer II, J. A Haven for Hate: The Foreign and Domestic Implications of Protecting Internet Hate Speech under the First Amendment. *Southern California Law Review*. 75(6), 2002, pp. 1493-1528
- [19] Buss, A. *The psychology of aggression*. Hboken, NJ: John Wiley & Sons, Inc, 1961.
- [20] Carré, J.M., Ruddick, E.L., Moreau, B.J.P. and Bird, B.M. Testosterone and Human Aggression. In *The Wiley Handbook of Violence and Aggression*, P. Sturmey (Ed.), 2017.
- [21] *Chaplinsky v. New Hampshire*, 315 U.S. 568 (1942)
- [22] Cherek, D. R. Effects of smoking different doses of nicotine on human aggressive behavior. *Psychopharmacology*, 75(4), 1981, pp. 339–345.
- [23] Cline, V. B., Croft, R. G., & Courier, S. Desensitization of children to television violence. *Journal of Personality and Social Psychology*, 27(3), 1973, pp. 360–365.
- [24] CNN. Quebec mosque suspect known to people who monitor far right groups. <https://www.cnn.com/2017/01/31/americas/quebec-mosque-shooting-suspect/index.html>, 2017.
- [25] CNN. El Paso mass shooting is at least the third atrocity linked to 8chan this year. <https://www.cnn.com/2019/08/04/business/el-paso-shooting-8chan-biz/index.html>, 2019.
- [26] *Cohen v. California*, 403 U.S. 15 (1971)
- [27] Crovitz, H. F., Daves W. Tendencies to eye movement and perceptual accuracy. *J. Exp. Psychol.* 63, 1962, pp. 495–98.
- [28] Dawson, M. E., Schell, A. M., & Filion, D. L. The electrodermal system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* Cambridge University Press, 2007, pp. 159–181.
- [29] DeWall, C. N., Finkel, E. J., Lambert, N. M., Slotter, E. B., Bodenhausen, G. V., Pond, R. S., Renzetti, C M., Fincham, F. D. The Voodoo doll task: Introducing and validating a novel method for studying aggressive inclinations. *Aggressive Behavior* 39, 2013, pp. 419–439.
- [30] DeWall, C. N., Twenge, J. M., Gitter, S. A. and Baumeister, R. F. It's the thought that counts: The role of hostile cognition in shaping aggressive responses to social exclusion. *Journal of Personality and Social Psychology* 96, 2009, pp. 45–59.
- [31] Dodge, K. A., & Crick, N. R. Social information-processing bases of aggressive behavior in children. *Personality and Social Psychology Bulletin*, 16(1), 1990, pp. 8–22.
- [32] Ekman, P. Are there basic emotions? *Psychological Review*, 99(3), 1992, pp. 550-553
- [33] Epstein, S., & Taylor, S. Instigation to aggression as a function of degree of defeat and perceived aggressive intent of the opponent. *Journal of Personality*, 35, 1967, pp. 265–289.
- [34] Finkel, E. J., DeWall, C. N., Slotter, E. B., Oaten, M. and Foshee, V. A. Self-regulatory failure and intimate partner violence perpetration. *Journal of Personality and Social Psychology* 97, 2009, pp. 483–499.
- [35] Gendron, M., & Barrett, L. F. Reconstructing the Past: A Century of Ideas About Emotion in Psychology. *Emotion review: journal of the*

International Society for Research on Emotion, 1(4), 2009, pp. 316–339.

- [36] Gøtzsche-Astrup, O. The time for causal designs: Review and evaluation of empirical support for mechanisms of political radicalisation. *Aggression and Violent Behavior*, 39, 2018, pp. 90–99.
- [37] Harmon-Jones, E., & Harmon-Jones, C. Neural foundations of motivational orientations. In G. H. E. Gendolla, M. Tops, & S. L. Koole (Eds.), *Handbook of biobehavioral approaches to self-regulation*. Springer Science + Business Media, 2015, pp. 175–187.
- [38] Hogg, M. A., & Adelman, J. Uncertainty-identity theory: Extreme groups, radical behavior, and authoritarian leadership. *Journal of Social Issues*, 69(3), 2013, pp. 436–454.
- [39] Huesmann, L. R. . The role of social information processing and cognitive schema in the acquisition and maintenance of habitual aggressive behavior. In R. G. Geen & E. Donnerstein (Eds.), *Human Aggression: Theories, Research, and Implications for Policy*. New York: Academic Press, 1998, pp. 73-109.
- [40] Jakubowicz, A., Dunn, K., Paradies, Y., Mason, G., Bliuc, A.-M., Bahfen, N., Atie, R., Connelly, K. and Oboler, A. *Cyber Racism and Community Resilience: Strategies for Combating Online Race Hate*, Palgrave Macmillan, London. 2017.
- [41] Kay, C. A., and Eibach, R. P. Compensatory control and its implications for ideological extremism. *J. Soc. Issues* 69, 2013, pp. 564–585.
- [42] Klein, M. K., and Kruglanski, W. A. Commitment and extremism: a goal systemic analysis. *J. Soc. Issues* 69, 2013, pp. 419–435.
- [43] Levin, B. ‘Cyberhate: A Legal and Historical Analysis of Extremists’ Use of Computer Networks in America’, *American Behavioural Scientist* 45, no.6, 2002, pp. 958–88.
- [44] Lieberman, J D.; Solomon, S; Greenberg, J; McGregor, H A. A hot new way to measure aggression: Hot sauce allocation. *Aggressive Behavior*. 25 (5), 1999, pp. 331–348.
- [45] McCauley, C., & Moskalenko, S. Understanding political radicalization: The two pyramids model. *American Psychologist*, 72(3), 2017, pp. 205–216.
- [46] McGregor, I., Hayes, J., & Prentice, M. Motivation for aggressive religious radicalization: Goal regulation theory and a personality×threat×affordance hypothesis. *Frontiers in Psychology*, 6, 2015.
- [47] Medium. 2020. Terrorgram: A community built on hate. <https://medium.com/dfrlab/terrorgram-a-community-built-on-hate-e02fd59ee329>
- [48] Miller, E. K., & Buschman, T. J. Cortical circuits for the control of attention. *Current opinion in neurobiology*, 23(2), 2013, pp. 216–222.
- [49] Moghaddam, F. The staircase to terrorism; a psychological exploration. *Am. Psychol.* 60, 2005, pp. 161–169.
- [50] Moore, T., & Zirnsak, M. Neural mechanisms of selective visual attention. *Annual Review of Psychology*, 68, 2017, pp. 47–72.
- [51] Neumann, P. Perspectives on radicalisation and political violence: Papers from the first international conference on radicalisation and political violence, International centre for the study of radicalisation and political violence. London, 17–18 January, 2008.
- [52] Pederson, W. C., Vasquez, E. A., Bartholow, B. D., Grosvenor, M. and Truong, A. Are you insulting me? Exposure to alcohol primes increases aggression following ambiguous provocation. *Personality and Social Psychology Bulletin* 40, 2014, pp. 1037–1049.
- [53] *R.A.V. v. City of St. Paul*, 505 U.S. 377 (1992)
- [54] *Reno v. American Civil Liberties Union*, 521 U.S. 844 (1997)
- [55] Sageman, M. *Leaderless Jihad: Terror Networks in the Twenty-First Century*. Philadelphia: University of Pennsylvania Press, 2008.
- [56] Saleem, M., Anderson, C. A. and Gentile, D. A. Effects of prosocial, neutral, and violent video games on children’s helpful and hurtful behaviors. *Aggressive Behavior* 38, 2012, pp. 281–287.
- [57] Silber, D. M. and Bhatt, A. *Radicalization in the West: The Homegrown Threat*. New York: NYPD Intelligence Division, 2007.
- [58] Silke, A. Holy warriors: Exploring the psychological processes of jihadi radicalization. *European Journal of Criminology*, 5(1), 2008, pp. 99–123.
- [59] Stephan, W. G., and Stephan, C. W. “An integrated threat theory of prejudice” in *Reducing prejudice and discrimination*. ed. S. Oskamp (Hillsdale, NJ: Lawrence Erlbaum), 2000, pp. 23–46.
- [60] Sternberg, Robert. *A Duplex Theory of Hate: Development and Application to Terrorism*,

Massacres, and Genocide. Review of General Psychology. 7. 2003, pp. 299-328.

- [61] Strack, F., & Deutsch, R. Reflective and Impulsive Determinants of Social Behavior. *Personality and Social Psychology Review*, 8(3), 2004, pp. 220–247.
- [62] Taylor, S. P. "Aggressive behavior and physiological arousal as a function of provocation and the tendency to inhibit aggression1". *Journal of Personality*. 35 (2), 1967, pp. 297–310.
- [63] Tedeschi JT, Felson RB. *Violence, Aggression, & Coercive Actions*. Washington, DC: Am. Psychol. Assoc, 1994.
- [64] UNOHCHR. International Convention on the Elimination of All Forms of Racial Discrimination. <https://www.ohchr.org/EN/ProfessionalInterest/Pages/CERD.aspx>, 1965.
- [65] Van Stekelenburg, J. Radicalization and Violent Emotions. *PS: Political Science & Politics*, 50(4), 2017, pp. 936-939.
- [66] VOX-Pol. The online life of a modern terrorist: Anders Behring Breivik's use of the internet. <https://www.voxpol.eu/the-online-life-of-a-modern-terrorist-anders-behring-breiviks-use-of-the-internet/>, 2014.
- [67] Wade, N.J., How were eye movements recorded before Yarbus? *Perception*, 44, 2015, pp. 851–883
- [68] *Watts v. United States* 394 U.S. 705 (1969)
- [69] Webber, D., & Kruglanski, A. W. The social psychological makings of a terrorist. *Current Opinion in Psychology*, 19, 2018, pp. 131–134.
- [70] Wiktorowicz, Q. *Radical Islam Rising: Muslim Extremism in the West*. Oxford: Rowman & Littlefield, 2005, pp. 77–78.
- [71] Zillmann D. Cognition-excitation interdependencies in aggressive behavior. *Aggress. Behav.* 14, 1988, pp. 51–64.