

University of Arkansas, Fayetteville
ScholarWorks@UARK

Theses and Dissertations

5-2020

Calling Out the Trolls: Responses to Witnessing Use of the “Troll” Label as a Defense in an Online Group Context

Dene E. M. Wamsley
University of Arkansas, Fayetteville

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



Part of the [Behavior and Behavior Mechanisms Commons](#), [Experimental Analysis of Behavior Commons](#), [Psychological Phenomena and Processes Commons](#), and the [Social Media Commons](#)

Citation

Wamsley, D. E. (2020). Calling Out the Trolls: Responses to Witnessing Use of the “Troll” Label as a Defense in an Online Group Context. *Theses and Dissertations* Retrieved from <https://scholarworks.uark.edu/etd/3570>

This Thesis 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.

Calling Out the Trolls: Responses to Witnessing Use of the “Troll” Label as a Defense in an
Online Group Context

A thesis submitted in partial fulfillment
of the requirements of
Master of Arts in Psychology

by

Dené E. M. Wamsley
University of Arkansas
Bachelor of Science in Psychology, 2016

May 2020
University of Arkansas

This thesis is approved for recommendation to the Graduate Council.

Denise Beike, Ph.D.
Thesis Director

William Levine, Ph.D.
Committee Member

Scott Eidelman, Ph.D.
Committee Member

Abstract

Although the term “troll” has existed since the 1980s, its meaning has shifted in recent years as social media use has increased. People provide contrasting and imprecise definitions for what constitutes “trolling,” and often apply the term subjectively to describe online discussants who are uncivil, who are deviant, and who and present counter-attitudinal opinions. Exposure to deviance, counter-attitudinal information, and incivility often leads to unwanted psychological effects. In theory, labeling an uncivil, counter-attitudinal deviant as a “troll” proposes that their intention is to disrupt the conversation and upset other discussants, which provides a reason for why incivility is used, and diminishes the threat of counter-attitudinal exposure. Participants were placed into an online discussion about transgender identities; while participants believed they were discussing with real people, they were actually interacting with a scripted computer program. All discussion comments were pro-attitudinal, save one. A 2 X 3 between-subjects design was employed to examine the effects of the civility of the counter-attitudinal comment (civil vs. uncivil) and the label used against this counter-attitudinal discussant (no label vs. rude label vs. troll label). Incivility exposure overall produced higher ratings of anger, attitude certainty, intentions to participate, and identification with the discussion group. The rude label overall decreased attitude certainty, while the troll label overall increased identification with being a person with their pre-existing attitude. In the uncivil condition, participants were marginally more willing to participate again when the troll label was applied, when compared to the two other label conditions. The intersecting influences of gender, pre-existing attitudes, and suspicions about the deception used are discussed.

Table of Contents

Introduction.....	1
Method.....	32
Results.....	35
Discussion.....	42
References.....	59
Tables.....	69
Figures.....	83
IRB Letter.....	96
Questionnaires.....	97
Script.....	99

Calling Out the Trolls: Responses to Witnessing Use of the “Troll” Label as a Defense in an Online Group Context

Introduction

A savvy internet user is likely familiar with the age-old advice to “ignore the trolls” – but what exactly constitutes as trolling? The concept has existed since early computer use in the late 1980’s, arising primarily on the message board Usenet (Herring, Job-Sluder, Scheckler, & Barab, 2002). At its origin, the term “trolling” was used to describe the very specific behavior in which an individual disrupts an online discussion by deceiving other users. This deception often involves pretending to not understand certain elements of the online forum, or pretending to advocate for an inflammatory, outrageous, and offensive stance (Schwartz, 2008). Originally, this behavior was intended to be humorous – it is funny if the troll can “bait” an individual into a pointless argument. This sense of humor hinges on the idea that the trolling behavior is *so absurd* that only a foolish, naïve individual would be hoodwinked into responding. This illustrates why the advice to “ignore the trolls” is so imperative – if you give them attention, you are the chump, but if the troll cannot bait anyone, he/she loses. While this was the predominant notion of trolling during the early stages of the internet, the meaning of the term, like most elements of language, has shifted as internet use has become more prevalent.

As the age of technology has propelled individuals into more frequent computer-mediated communication (CMC), average individuals have become more familiar with the idea of trolling – even less proficient internet users are likely aware of the concept. Rather than the specific term it once was, trolling has come to denote a variety of behavior, from light-hearted and humorous to incredibly vitriolic. The various meanings that the term “trolling” can encompass presents a particular puzzle: when someone labels another user a “troll,” what exactly

do they mean by it, and what motivates them to use this specific term? The frequency with which individuals online are accusing other discussants of being “trolls” has increased rapidly, and therefore it is imperative to understand the function of such a term in the context of an online discussion. Investigating this phenomenon through the lens of social psychology can elucidate the effects of using such a term to describe people and their behavior in a social setting.

An open-ended survey of 955 participants from an undergraduate sample illustrated the various ways in which people define the concept of trolling (Wamsley, 2018). Only 9.5% of respondents considered trolling to be a humorous behavior, giving definitions like “trolling is when you post meaningless or sarcastic/humorous comments that don't apply to the original post” and “trying to be funny for the readers, and gain as many likes as possible.”. On the other hand, 81% considered trolling to be blatantly destructive and insidious, giving definitions like “[harassing] or bothering people for no reason other [than] personal enjoyment” and “trolling is when you make an online post with the deliberate intention of harming, angering, or irritating another person”. Only 7.8% of respondents gave definitions of trolling behavior that involved deception or trickery, which mirrors the original definition of trolling from the days of early CMC. A quarter of respondents specifically compared trolling with bullying, cyber-bullying, harassment, and stalking; the overt intention of the troll posited by these definitions starkly differs from the covert intention suggested by the original definition. Together, the countless proposed definitions and intentions of trolling behavior create a concept that is nebulous, complicated, and imprecise in nature, and should therefore be further investigated.

Additionally, an empirical study was conducted to explore under which parameters the “troll” label is more likely to be applied (Wamsley, 2018). An undergraduate sample of 340 students from the University of Arkansas were provided with eight different surreptitiously

fictitious comments on articles posted to Facebook. The articles and comments involved President Donald Trump and the current NFL kneeling protests. Comments varied on their political affiliation, either conservative or liberal in nature, and varied in their civility, either mild or inflammatory. Respondents rated each comment on how rude/inconsiderate it was, how funny it was, and how likely they thought it was that the comment was trolling. Participants additionally provided their own political affiliation in order to analyze their reactions as pro-attitudinal and counter-attitudinal.

Overall, inflammatory comments and counter-attitudinal comments were rated as rude/inconsiderate and as trolling higher than their mild and pro-attitudinal counterparts. While uncivil comments were rated as funnier than mild comments overall, counter-attitudinal posts were rated as less funny than pro-attitudinal comments. Furthermore, uncivil counter-attitudinal comments and uncivil pro-attitudinal comments were rated as trolling at similar levels; there was no significant difference between these conditions. However, mild counter-attitudinal comments were rated as trolling at a significantly higher rate than mild pro-attitudinal comments. This suggests that when respondents witness incivility, the content of the comment (pro- or counter-attitudinal in nature) does not necessarily matter. People are willing to label an uncivil, pro-attitudinal comment as trolling, even when they agree with its substance. Alternatively, when a political comment is mild in nature, the troll label is significantly more likely to be applied to a comment that is counter-attitudinal to the respondent's own political stance. These results pose yet another empirical question to be examined: if the "trolling" label is more likely to be applied to uncivil and counter-attitudinal online posts, does the label function as a defense to incivility and counter-attitudinal exposure? In other words, are people motivated to use this term in order to reduce the potential negative effects of such exposure?

To begin to answer this question, it is important to consider why “trolling” is such a subjective label. When interacting with other entities online, many cues that typically help interpret the conversation become unavailable; for example, voice intonation and facial expressions are virtually absent (Walther & Parks, 2002). This makes inferring an online speaker’s intention rather difficult. With the ever-increasing use of social networking sites (SNS) like Facebook – in which users can interact, communicate, and discuss any topic of interest – it is rather common to witness a user accusing another user of being a “troll.” Often, this accusation is observed when the target user’s posts are uncivil or counter-attitudinal to the accuser’s beliefs. Moreover, it often occurs when users are discussing topics such as politics, religion, or issues of social justice – all of which regularly suffer from a vast amount of incivility use (Coe, Kenski, & Rains, 2014).

When an individual considers labeling another user as a “troll,” they must infer the motivations of said user. Unfortunately, the motivation that is inferred and the target’s real motivation frequently differ; this disparity can lead to misunderstandings or inaccurate labels (Hardaker, 2010). Because of this, the act of labeling another user as a troll is a subjective, inferential behavior that can lead to an incorrect accusation.

The act of labeling of an online user as a troll has very seldom been investigated. In today’s digital age in which users discuss substantial topics in an online context, it is crucial to understand when and why the troll label is applied. As illustrated, this label is more frequently used against a user that presents an uncivil, counter-attitudinal opinion. Thus, it is important to analyze the possible effects of witnessing an uncivil individual you disagree with being labeled as a troll. Labeling a user as a troll suggests that their opinion is motivated by at least one of three common motivations associated with the concept of trolling: the desire to upset other

discussants for the sake of stirring the pot, the purpose to be humorous, or the goal to surreptitiously take on an offensive stance that they *don't truly believe* to cause argumentation. In contrast, a genuine (non-troll) actor who presents an uncivil, counter-attitudinal opinion may have the intention to change peoples' minds, reach an accurate conclusion, and/or assert an opinion they believe to be sincerely factual and vital. One may be considered "rude" if they offer an uncivil, counter-attitudinal opinion, but this does not mean that they do not truly believe in their point of view and desire others to thoughtfully deliberate on their opinion.

The former assumption is far less threatening to the individual – if an uncivil, counter-attitudinal user is simply a troll, then their opinion does not have to be sincerely considered. The opinion becomes easier to cognitively discount. Additionally, the sage advice to "ignore the trolls" compels users who label others as trolls to overlook them – if you ignore the trolls, they become harmless. The very act of engaging with a user they consider to be a troll would theoretically give the troll satisfaction. Moreover, someone who is a troll is a non-credible source, and therefore their argument should never be considered. On the other hand, if a discussant is genuine and wants to have a fruitful debate, there would be no harm in responding to their counter-attitudinal opinion in a thoughtful manner. In the context of an online group, if one in-group member labels an out-group member as a troll, they are signaling to other members that this out-group member's opinion should not be thoroughly considered or replied to. Therefore, the out-group member's opinion becomes less threatening to group members overall. This further suggests that the troll label gives discussants a way to disregard an opinion that threatens the group and their closely-held beliefs.

This specific phenomenon has not yet been examined by extant research. Therefore, the primary goal of this thesis is to determine the effects of witnessing the "troll" label being applied

to an online discussant on various outcomes, and how the nature and implications of the label potentially vary from other ad-hominem labels, such as “rude”. While calling an uncivil, counter-attitudinal discussant “rude” is a method of derogation, this label fundamentally differs from the “troll” label in its assumption of the target’s motivation. More specifically, I argue that perceiving a counter-attitudinal discussant as a “troll” functions as a defense against incivility and incidental exposure to counter-attitudinal information, as well as a form of deviance derogation.

Existing Trolling Literature

Due to the rising prevalence of SNS, researchers have turned their focus toward understanding and defining the darker aspects of online interactions, including trolling. However, researchers differ on what specific behavior they consider to be trolling, which complicates the literature. Often, these studies conflate trolling with bullying, harassment, flaming, and cyber-crimes such as hacking (Cho & Kwon, 2015; Fichman & Sanfilippo, 2015; Shachaf & Hara, 2010). Whether these concepts and trolling are completely synonymous is still to be determined. There seems to be an element that differentiates the concept of trolling from other impolite online behavior. Hardaker (2010) suggests that trolling behavior is specifically *deceptive*, *aggressive*, *disruptive*, and *successful*. However, this distinction is not necessarily normative in the trolling literature, as some researchers occasionally operationalize a wide variety of discourteous behavior as trolling.

An impactful portion of the trolling literature is focused on the personality and motivations of the troll themselves. For example, a frequently cited article by Buckels, Trapnell, and Paulhaus (2014) demonstrated that participants who indicated enjoyment from trolling others online were also more likely to have higher scores for sadism, Machiavellianism, and

psychopathy. This finding is consistent with the concept that trolls desire conflict for the sake of enjoyment. Shachaf and Hara (2010) also argued that Wikipedia “trolls” not only engaged in repetitively harmful behavior with destructive intent, but also gained enjoyment from this behavior. Ample subsequent research has continued this line of work by honing in on which personality characteristics and motivations influence the propensity to troll (Craker & March, 2016; March et al., 2017; Lopes & Yu, 2017; Seigfried-Spellar & Lankford, 2018; Sest & March, 2017; Thacker & Griffiths, 2012).

Trolling behavior involves two sides: the troll and the recipient(s). Social-psychological researchers are beginning to investigate both the perception of trolling and its impact on the audience. For instance, an early article by Herring and colleagues (2002) analyzed how an online feminist community struggled to combat a particularly noxious and effective troll. In the context of video gaming, self-report data suggests that being the target of trolling behavior takes a toll on self-esteem (Thacker & Griffiths, 2012). Similarly, research has shown that male participants are more likely to report the intention to engage/argue with a troll, while female participants were more likely to say they would ignore them; additionally, participants assumed that male trolls were motivated by malevolence and provocation more so than their female counterparts (Fichman & Sanfilippo, 2015). Research aiming to gauge what personality characteristics people assign to a troll found that participants largely responded that trolls were attention-seeking, conflict-seeking, funny, annoying, and vicious. However, when participants were asked to view fictitious profiles and rate how likely it was that the profile user was a troll, participants applied the troll label only to profiles that suggested attention/conflict-seeking and viciousness; funny profiles were *not* highly rated as trolls. (Maltby et al., 2016).

The research outlined so far suggests that the application and meaning of the “troll” label can fluctuate between individuals and contexts. This is substantiated by research by Coles and West (2016), who argue that the meaning of the word “troll” is determined through interaction and conversation by individuals, rather than by the dictionary or authority figures such as psychological researchers. Through a data-analysis of 107 pre-existing online forum comments concerning the concept of trolling, they found that online discussants believe that trolls are easy to identify, that trolling behavior has shifted over the past years to become less sophisticated, that trolls are “nasty”, and that the online community should counteract trolls as a form of “vigilantism” (Coles & West, 2016).

As of today, the literature examining trolling has been largely focused on personality traits and motivation of the troll themselves. While research investigating the impact on and perception by online audience members is becoming more frequent, it is crucial to continue to explore this side of the story.

Review of Relevant Classic Literature

The act of labeling another individual as a “troll” can most frequently be witnessed in an online, social networking context. However, the phenomena that I theorize are related to this process have been long studied in face-to-face contexts as well. As future social-psychological research takes strides towards studying human behavior through CMC, it is important to consider analogous processes that do not happen behind a screen. The following section reviews and summarizes the long-standing psychological literature on in-person human interaction in the following subcategories of research.

Functions of Labeling

The human brain is frequently motivated to reduce cognitive effort spent; often, people take cognitive short cuts in order to preserve mental exertion. One such method that saves cognitive energy is the process of putting people and objects into *categories*. A common way by which people can indicate which categories they consider people/objects to fall under is by assigning them a *lexical label*.

A large amount of cognitive literature concerning categorical labeling of others involves the perspective of those being labeled and third-party perceivers of the labeled. More specifically, the impact of using of stigma-related labels involving sexual identity and mental health has been frequently investigated (e.g. Fasoli et al., 2016; Foster, Schmidt, & Sabatino, 1976; Freeman & Algozzine, 1980; Phelan et al., 2018). Similarly, the impactful *Labeling Theory* suggests that labeling an individual as “deviant” may have an unintended rebound effect (Becker, 1963; Lemert, 1951). This line of research is primarily focused on the influence of labels on only one side. However, as with the concept of trolling, labeling involves two sides: the labeler and the label recipient. What is the cognitive function of such categorical labels for the labeler?

Social and cognitive psychology intersect in their investigation into social-categorical boundaries. In many cases, the act of labeling another individual denotes group membership, what social norms become salient, and how he/she should be treated. Once this categorical label has been assigned to a person, we tend to make inferences about both the specific individual, as well as other individuals in the same category; these inferences impact our responses to categorical exemplars (Bruner, Goodnow, & Austin, 1956; Lau et al., 2018; Tajfel & Turner, 1979). Categories that have been studied through this lens are often related to racial and gender

constructs (Chen et al., 2018; Chen & Ratliff, 2015; Gelman, Collman, & Maccoby, 1986; Kleider-Offutt et al., 2018; Ito & Urland, 2003; Maccoby, 1989).

Extant research has not yet considered the impact of using the “troll” label to denote categorical membership. According to participants’ definitions of “trolling,” people who fall under this category are likely seeking conflict, trying to be funny, or pretending to believe in a provocative stance to upset others (Maltby et al., 2016). It has been illustrated that indicating category membership of a person can influence how we interact with that person in the future; therefore, it is likely that the act of labeling someone a “troll” and signifying his/her membership to the “troll” category may influence perceivers to “ignore the troll”, to not thoughtfully deliberate on the “troll’s” posts, and to cognitively discount them. Since the label is often used in the face of incivility and counter-attitudinal opinion, it is further important to consider how indicating membership to the “troll” category may influence how people perceive and defend against these threats.

Tangentially, the “troll” label can be considered an *ad-hominem* attack against other discussants. Rather than attacking the substance of their argument, using this label attacks their character. Although the use of ad hominem attacks is considered a logical fallacy, it is frequently employed during online arguments. This suggests that people are highly motivated to denounce the character of a counter-attitudinal discussant. According to van Eemeren and colleagues, the use of an ad hominem attack is an attempt to discredit the opponent, make their opinion obsolete, and end the discussion (2012). Rather than coming to an agreement or thoughtful conclusion, the ad hominem attack ends the discussion in a more abrupt way.

While calling an opponent “rude” is also considered an ad hominem attack, it does not function to undercut the opponent’s authority as viciously as the label “troll” does. Someone who

is a “troll” has ill intent – whether that is to harm others, disrupt and deceive, or to be funny – and does not wish to have a thoughtful and deliberative discussion. Therefore, calling a counter-attitudinal discussant a “troll” is a rather extreme form of ad hominem that likely is used as an attempt to curtail the discussion.

Defense against Counter-Attitudinal Exposure

Every day, we are exposed to opinions and attitudes that starkly differ from our own. In some cases, we are inclined to listen and learn such information, despite its contradiction to our pre-existing bias. However, there are several processes that can occur that reduce the impact of counter-attitudinal information on our deliberation, which serve as a defense against it. In other words, rather than learning and processing a counter-attitudinal opinion, individuals are often motivated to discount it, argue against it, and sometimes flat out ignore it. Such processes may be primarily advantageous, but also reduce the diversity of knowledge and opinion within the individual.

For example, a robust process that functions to make counter-attitudinal information less threatening to one’s own prior beliefs is *disconfirmation bias*. The disconfirmation bias model suggests that, in the face of a counter-attitudinal argument, individuals will call on their prior, incompatible beliefs in order to more harshly scrutinize the information at hand (Edwards & Smith, 1996). A component to this model, dubbed the *prior belief effect*, suggests that individuals often become more polarized in their prior point of view when exposed to a belief that they disagree with, because they consider the counter-attitudinal argument to be inherently less convincing (Lord, Ross, & Lepper, 1979).

Such effortful scrutiny of counter-attitudinal information can result in hostile responses, including arguing against the information, as well as belittling it (Taber & Lodge, 2006). As

opposed to a deliberative response in which the effortful inspection leads to understanding and integration with the individual's prior beliefs, hostile responses tend to increase polarization and prevent positive outcomes, such as open-mindedness. In fact, witnessing counter-attitudinal information being explained away can activate the reward center in the brain (Westen et al., 2006). If there are potentially rewarding outcomes for dismissing a counter-attitudinal argument, this suggests that individuals may be motivated to do so.

Similarly, the theory of *motivated reasoning* posits that individuals can have one of two goals when contemplating an argument: reaching an accurate conclusion (accuracy goals), or reaching a biased, directional conclusion (directional goals) (Kunda, 1990). Accuracy goals can lead to increased effort and contemplation of the information at hand with the epistemological intention to fully understand and accurately represent it; additionally, individuals operating with accuracy goals, as opposed to directional goals, are less likely to rely on primacy effects and stereotypic information, as well as less likely to come to extreme, polarized conclusions (Kruglanski & Freund, 1983; Tetlock, 1985). On the other hand, Kunda argues that an individual with a direction goal desires to construct a rational, persuasive argument in order to convince a counter-attitudinal discussant, rather than to find an accurate conclusion (1990). While the individual feels he/she is being impartial, this is often not the case (Pyszczynski & Greenberg, 1987). These biased motivations can impact what evidence and justifications are called upon to bolster the directional conclusion.

A related line of research on *dissonance theory* suggests that people are so uncomfortable with inconsistency that they will adapt their attitudes to match a freely-engaged in counter-attitudinal behavior (Cooper & Fazio, 1984; Festinger, 1957; Festinger & Carlsmith, 1959). In relation to motivated reasoning theory, it is likely that reaching an accurate conclusion that

contrasts a pre-existing belief induces feelings of dissonance; therefore, people may feel more comfortable approaching the argument with directional goals (Kunda, 1990).

Finally, *self-affirmation theory* is another approach that aims to explain people's underlying mechanisms in the face of counter-attitudinal information. Steele posits that an individual is motivated to maintain the overall notion that he/she is a good, virtuous person (1988). Strongly-held opinions and attitudes are often relevant to a person's self-concept; in the face of counter-attitudinal or dissonance-inducing information, people may feel threat to their self-concept. Abundant research has suggested that people are less threatened by a counter-attitudinal attitude in one domain when they are able to enhance their self-concept in a different domain (Sherman & Cohen, 2006). If a person cannot bolster their self-identity in a different domain, it is likely that they will continue to defend against the counter-attitudinal threat.

Overall, previous psychological research has demonstrated multiple possible explanations for why people are motivated to reduce the impact of a counter-attitudinal argument. This phenomenon is fundamentally crucial to consider, since the "troll" label is often applied to discussants who present ideologically-incongruent opinions to the labeler's own beliefs.

Face-to-Face Incivility

Whether with a co-worker, a classmate, a family member, a politician, or even with a stranger, face-to-face uncivil discussion can be incredibly difficult, uncomfortable, and psychologically taxing for the parties involved, as well as third-person spectators.

Prior research suggests that the presence of incivility in such contexts can lead to a multitude of harmful effects, including workplace "burnout" (Liu et al., 2018), stress and negative affect (Webster et al., 2018), emotional exhaustion (De Clercq et al., 2018), feelings of exclusion and decreased engagement (Schilpzand & Huang, 2018), school dissatisfaction and

damage to class performance (Caza & Cortina, 2007), job dissatisfaction and increased counter-productive work behavior (Cortina et al., 2001; Penney & Spector, 2005), and a decline in overall employee well-being (Lim & Cortina, 2005).

While these in-person effects are important for consideration, comparable effects have been observed on the internet as well. The open access to a vast array of different people across the globe as well as the deliberative potential of social networking sites allows for online users to have deep, meaningful discussions. However, these discussions frequently turn uncivil – thus it is additionally important to understand the outcomes of incivility in an online context. Often, online incivility differs in that people are dissociated from the people they are discussing with; additionally, people can often opt in and out of the discussion at will, as opposed to being unable to leave a workplace or classroom that suffers from incivility use. Such outcomes and nuances are unpacked subsequently.

Group Deviance

Classic social-psychological theory has long been interested in detailing the phenomenon of group deviance – the perceived violation of a group norm. Group norms can be set by formally stating rules and regulations (including expectations of civility during group member interactions). Conversely, groups are often held together by perceived similarity; therefore, expressing dissimilarity to the group's identity may be perceived as group deviance. Group members react to such deviance in a multitude of ways (Levine, 1989).

Early research by Festinger (1950) suggests that informal groups are commonly held together by a homogeneity of attitudes. The *social identity perspective* conceptualizes of groups in a similar manner. This perspective theorizes that an individual's self-concept is an amalgamation of several components, including which social groups or categories they consider

themselves to belong to (Hogg & Reid, 2006; Tajfel & Turner, 1979; Turner et al., 1987). A subset of this perspective investigates how people cognitively represent groups as categories – people are inclined to over-emphasize similarities of groups they belong to, and differences between the groups they belong to and those they do not (Tajfel, 1959). This creates a clear delineation between group boundaries that helps to preserve cognitive energy in the future.

Through this perspective, it is clear why expression of dissimilarity by a group member is often perceived as deviance. Similarity between group members is frequently the glue that keeps them together. Informal groups form for a myriad of reasons. For example, you may choose to join a group of advocates for playing video games, because your hobby of playing video games is highly relevant to your self-concept, and you enjoy discussing it with like-minded individuals. The enjoyment of playing video games becomes the expressed norm of the group. If a member of this group suddenly states that “video games are a waste of time”, this hinders the group’s harmony and homogeneity, and expressly contrasts the ideal of the group’s identity. For this reason, expression of an opinion that is counter-attitudinal to the group’s identity can be considered deviance, as well, and often produces strong responding (Abrams & Hogg, 1990).

Based on the social identity theory, the *Black Sheep Effect* posits that ratings of in-group members are more extreme than out-group members; more specifically, in-group members that violate an in-group behavioral norm are judged more severely than norm-violating out-group members (Marques, Yzerbyt, & Leyens, 1988). In other words, in-group members’ behavior is put under more scrutiny and expectation, because their behavior reflects the identity of the group overall.

Several empirical studies have targeted the underlying mechanisms, impacting factors, and various responses involving the Black Sheep Effect. For instance, when a group member

strongly self-identifies with the group, he/she considers an in-group deviant to be less typical of the group overall, which may function to make the deviance less threatening to the group's positive stereotype (Castano et al., 2002). This becomes harder to do when individuals are put under cognitive load, which suggests an underlying, effortful cognitive process (Yzerbyt, Coull, & Rocher, 1999).

While harsher derogation of a norm-violating in-group member may serve to preserve the group's identity, the social identity perspective suggests that an individual's own self-identity is intrinsically tied to that of the group. Therefore, a deviant in-group member additionally threatens the individual self-concept of group members themselves. Eidelman and Biernat argued that in-group members may react to perceived deviance at an individualistic level in order to protect their own personal identities (2003). Participants in this paradigm were given the opportunity to derogate the unfavorable deviant, or to disidentify with the group that experienced deviance. The order of these options varied; when participants could disidentify with the group first, they were less likely to derogate the target, and vice versa. This suggests that these two options can substitute one another, and therefore both act as preservation of one's self identification in the face of an unfavorable group member (Eidelman & Biernat, 2003).

Does derogation function to exclude the deviate, or is the goal to re-assert the group's norms? Eidelman and colleagues further investigated this question in 2006. Participants were provided with the opportunity to exclude in-group deviants from the group boundaries, and to devalue the deviant. The order of these options varied once more; despite the counterbalanced order, participants consistently excluded the in-group deviant from the group boundary. However, when exclusion occurred first, participants no longer differentially devalued the in-group deviant more than the out-group deviant (Eidelman, Silvia, & Biernat, 2006).

Overall, empirical research has suggested that group members are motivated to both exclude and derogate opinion-based deviants in order to maintain group homogeneity and prevent their own identity subversion.

Taking a CMC Approach

The use of the “troll” label against uncivil, counter-attitudinal discussants can most frequently be observed in *online* discussions. As computer-mediated communication is becoming more commonly utilized and accessible to the public, it is vital to inspect whether frequently tested in-person psychological phenomena function in a similar manner in an online context. Therefore, the following sections outline how specific psychological research is advancing our understanding of how key factors manifest in a CMC context.

Online Counter-Attitudinal Exposure

A key component of scenarios in which the troll label is used is that of counter-attitudinal exposure. Social networking sites (SNS) provide ample opportunities to express and discuss diverse attitudes on various topics. Regarding political opinion, even online spaces intended to discuss irrelevant topics can be plagued with the presence of political discussion (Wojcieszak & Mutz, 2009). Therefore, coming across a counter-attitudinal political argument online is nearly inevitable (Garrett, Carnahan, & Lynch, 2013). The cognitive and behavioral processes involved when exposed to counter-attitudinal information – which have been thoroughly illustrated by social-psychologists both in face-to-face contexts and online contexts – are crucial for consideration since the “troll” label is frequently used against ideological opponents online.

The deliberative potential and the accessibility of political content of SNS makes them fertile ground for debate, and therefore counter-attitudinal exposure (Brundidge, 2010). Ample research is concerned specifically with the act of online *selective* exposure (Dylko et al., 2017;

Garrett & Stroud, 2014; Knobloch-Westerwick, Johnson, & Westerwick, 2014; Weeks et al., 2017; Westerwick, Johnson, & Knobloch-Westerwick, 2017; Yang & Unnava, 2016). This concept can be considered the combination of selective approach, in which individuals systematically seek out pro-attitudinal information, and selective avoidance, in which they routinely avoid counter-attitudinal information (Garrett & Stroud, 2014). While both approach and avoidance occur, research suggests that selective approach online is a stronger influence when selecting news articles for consumption (Garrett & Stroud, 2014).

This bias to confirm pre-existing beliefs through selective exposure drives individuals' online behavior in numerous ways. Apart from selective exposure, however, *incidental* exposure to counter-attitudinal information also occurs. Incidental exposure consists of accidentally “stumbling” across information, as opposed to purposefully seeking it out (Lee & Kim, 2017). Regarding the “trolling” label, this phenomenon is likely related. An individual may purposefully seek out a comment section in which most commenters are pro-attitudinal in opinion; however, since comment sections are not often thoroughly moderated, the individual will likely encounter a counter-attitudinal commenter as well. This scenario is when the “troll” accusation most frequently occurs.

Incidental exposure may be particularly dissonance-inducing – the exposure is not intentional, and therefore the person who comes across counter-attitudinal information is likely not prepared to digest it (Weeks et al., 2017). Therefore, incidental exposure may provoke strong responding, as individuals are highly motivated to reduce dissonance (Festinger, 1957).

In rare cases, online incidental exposure leads to learning, in which the information is actually considered, and subsequently can be recalled; however, for information to be learned, the incidental exposure must be followed up with purposeful exposure, in which the individual

continues to seek out and read related information (Lee & Kim, 2016). In other words, simply seeing information accidentally and then ignoring or cognitively discounting it does not lead to remembering, let alone understanding or agreement.

Research also suggests that online incidental news exposure motivates political engagement and corrective action. A comprehensive survey suggested that incidental exposure to news online was positively related to both offline and online political participation (Kim, Chen, & Gil de Zúñiga, 2013). However, the attitudinal valence of such news was not specified or manipulated in this research. Research that specifically focused on counter-attitudinal news exposure determined that individuals believe said news to be inherently hostile and non-credible in nature; this perception of bias and hostility increases feelings of anger, which in turn increase the tendency to engage in political action online, such as publicly sharing pro-attitudinal sources (Hwang et al., 2008). Through this lens, the act of labeling someone a “troll” may function as a method of corrective action in the face of counter-attitudinal opinion. If one alerts his/her fellow pro-attitudinal commenters to the presence of a counter-attitudinal troll, then he/she is doing his/her fair share of corrective, political action.

Overall, the phenomenon of online incidental exposure and subsequent defense mechanisms against counter-attitudinal exposure are critical to understanding the function of the “troll” label, as these are the circumstances in which the label is most commonly applied. In order to strive for ecological validity, incidental counter-attitudinal exposure is a key component to the present research.

Online Incivility

A recurring theme for those who are labeled “trolls” in online discussions is the presence of incivility. In today’s climate, politics and incivility appear to go hand in hand. The use of

incivility by television news programs and newspaper sources has climbed sharply since the 20th century (Sobieraj & Berry, 2011). However, it is apparent that individuals are turning away from traditional sources such as television and are increasingly using social media for both social and political aspects (Kathurwar, 2017). Consequently, researchers have begun to focus on the prevalence of incivility on social networking sites (SNS). It is increasingly common to come across uncivil discussion on SNS such as Facebook and Twitter. Researchers have suggested that the role of anonymity has facilitated incivility use online (Borah, 2013; Graf, Erba, & Harn, 2017; Scott, Rains, & Haseki, 2011). However, users on websites such as Facebook operate profiles that are frequently connected to their personal information, their families, and even their occupations. Despite this, SNS like Facebook in which users are identified have become platforms on which uncivil discussions occur daily.

Comment sections and discussion threads on SNS are often a prime source for uncivil discussion, particularly those that discuss politics, foreign affairs, and economics (Coe, Kenski, & Rains, 2014). According to Pew Research, two-thirds of Americans self-report that they occasionally acquire political news through social media; however, over half of respondents are skeptical of the news they see (Matsa & Shearer, 2018). SNS has increasingly become a platform for news, and yet individuals are unsure of its credibility; this discrepancy creates the opportunity for further political debate and discussion. Unlike closed groups that individuals must choose to join, SNS comment sections and threads are commonly open to the public, have no formally stated rules, and tend to have virtually no moderation. Such open access to comment sections can provide an influential source of political deliberation, in which individual discussants can provide their insight and criticism towards a topic at hand (Manosevitch & Walker, 2002).

Exposure to uncivil debate in an online context has been demonstrated to elicit several different attitudinal responses from discussion participants. For example, the presence of incivility by unaffiliated commenters in the comment section of a news article decreases credibility ratings of the news source itself (Kim, 2015; Weber, Prochazka, & Schweiger, 2017). Fellow discussants themselves are more likely to be regarded as overly dominant and lacking in credibility when they exhibit uncivil tactics in the discussion (Ng & Detenber, 2006). While an optimistic goal of such discussion is to reach a deliberative agreement, empirical research demonstrates that uncivil debate exposure instead increases polarization of pre-existing attitudes and decreases open-mindedness (Anderson et al., 2014; Borah, 2014; Hwang, Kim, & Kim, 2018). Furthermore, incivility exposure increases perception of mass polarization (Hwang, Kim, & Huh, 2014). In other words, witnessing an uncivil discussion gives people the idea that most, if not all, discussions are also uncivil and polarized.

Although incivility exposure frequently leads to lower ratings of deliberative potential of the discussion (Hwang, Kim, & Huh, 2014) – which suggests that discussants do not believe the argument will have any productive consequences – online users are also likely to find such discussions more *entertaining* than their civil complements (Graf, Erba, & Harn, 2017; Mutz & Reeves, 2005). This could partially explain why individuals are also more likely to indicate desire to participate in uncivil discussions more so than civil discussions (Borah, 2014; Wang & Silva, 2018). This paints an interesting – albeit puzzling – picture: perhaps, some people seek out uncivil discussions not because they believe it will result in a constructive, deliberative conclusion, but because they find it entertaining.

Extant research on incivility exposure has additionally delved into its role on emotional responses. Such exposure can elicit certain negative emotions, which in turn may motivate

specific action tendencies. For example, Gervais (2017) demonstrated that exposure to an uncivil, counter-attitudinal argument elicited feelings of anger for participants; this anger, consequently, predicted an increase in incivility use within responses by the participants themselves. In a sense, counter-attitudinal incivility appears to be a vicious cycle that is influenced, in part, by augmented feelings of anger. While this cycle seems intuitive, similar research also found evidence that uncivil pro-attitudinal arguments bred more incivility use as well; however, anger no longer played a role (Gervais, 2015). Overall, exposure to uncivil argumentation appears to motivate additional use of incivility, but anger is only elicited when the argument is *counter-attitudinal* in nature.

Rösner and colleagues (2016) additionally sought to investigate the impact of civility exposure on negative affect and subsequent incivility use. They found that participants who were exposed to more incivility in a discussion reported more aggressive cognition. Counter to Gervais's findings, however, incrementally increasing the presence of incivility did *not* increase incivility use by participants (Rösner, Winter, & Krämer, 2016).

Analogous research further demonstrated that exposure to uncivil online discussion increased self-reported negative emotions (a conglomeration of anger, disgust, anxiety, worry, and fear) – these negative emotions, in turn, increased the propensity to participate in the discussion (Wang & Silva, 2018). However, the effect of these conditions changed for men and women, depending on the topic at hand; this suggests that the context/topic of the discussion is a key factor to consider when investigating emotional reactions.

While research on cyclical incivility use has produced contradictory results, it is apparent that negative affect can be elicited by the presence of counter-attitudinal incivility.

Operationalizations of Online Incivility

It is imperative to pinpoint what particular behaviors “incivility” encompasses. Early work by Papacharissi (2004) suggests that incivility and impoliteness are two distinct concepts: she specifically defines incivility as the obstruction of “democratic merit” in a discussion. She operationalizes incivility as threats to democracy and civil rights, and as the use of stereotypes (Papacharissi, 2004). However, Papacharissi does acknowledge that the decline of politeness within discussion is intrinsically tied to incivility.

Papacharissi’s (2004) operationalization of incivility is a departure from the norm within the incivility literature. Researchers more frequently consider incivility and impoliteness as virtually synonymous, which is typically how these two concepts are represented in public discourse. Other operationalizations of incivility include the following:

Insulting Language

Intuitively, insulting language involves the unnecessary use of offensive adjectives, often ad-hominem insults, and mockery that is intended to upset and condescend the target (Anderson et al., 2013; Borah, 2013; Gervais, 2015; Mutz & Reeves, 2005; Sobieraj & Berry, 2011; Thorson et al., 2010). Vulgar language such as profanity can also fall under this category (Coe, Kenski, & Rains, 2014).

A primary manipulation of this thesis involves varying what label is applied to a single counter-attitudinal poster: no label, “rude” label, or “troll” label. Calling a discussant “rude” can be considered an insulting adjective, which falls under the insulting language category. Therefore, the uncivil condition overall only utilizes language that insults political parties *as a whole*, and not single discussants, as to not introduce confounds with the intended manipulation.

Histrionic Language

Histrionic language involves displaying emotion through text (Gervais, 2015; Sobieraj & Berry, 2011). One form of emotional display is the use of capital letters and exclamation marks, which suggests a raised voice and excited expression. Additionally, histrionic language may include the employment of emotionally-charged words, such as “sad”, “fear”, and “angry” (Coe, Kenski, & Rains, 2014).

Ideological Misrepresentation & Exaggeration

This category of incivility includes using inflammatory hyperbole against a political party or standpoint with the goal to make it appear more radical (Coe, Kenski, & Rains, 2014; Gervais, 2015; Sobieraj & Berry, 2011). Like the strawman fallacy, this involves misrepresenting an agenda to make it easier to attack. For example, the suggestion that “feminists want to enslave men” is an ideological exaggeration about the goal of the feminist movement that makes it an easier target to vilify.

Responses to Online Group Deviance

In an online context, labeling another user a “troll” may be also a reaction to perceived *deviance*. Overall, online deviance can be considered and examined at two different levels: the macro, cyber-cultural level in which the norms of the overarching internet culture are deviated from, or the micro, CMC-processes level in which norms involved in communicating with other discussants in a specific group or setting are violated (Denegri-Knott & Taylor, 2005). The macro-level is often investigated by psychological research because online norms and standards have arguably shifted away from common offline norms. For example, illegal downloading and sharing of media such as music, movies, and even academic documents is committed by internet users at an alarming rate, which would suggest this form of online theft is not considered as

deviant as in-person theft (Helbig, 2014; Strauss, 2014). While certain discourteous behaviors such as trolling are more acceptable on specific online websites, such as 4chan, overall, people respond to perceived trolling in a negative manner (Maltby et al., 2016; Thacker & Griffiths, 2012). This suggests that while particular online groups may find these behaviors acceptable, other online spaces do not. Therefore, it is equally important to consider deviance at the micro-level.

Although comment sections – a common source of incivility, deliberation, the presence of counter-attitudinal opinion, and ultimately the accusation of trolling – are not defined specifically as a group with limitations to joining, they can represent a population with similar identities, and elicit strong feelings of ingroup inclusion (Mikal et al., 2015). For instance, research investigating comment sections on posts within the image-hosting website Imgur suggests that frequent Imgur users establish group norms through repeat comments, disapproving of deviance through the “down-voting” feature, and derogating the deviants through response comments (Mikal et al., 2014).

This can be extrapolated to SNS such as Facebook, particularly on politically-charged posts. For example, the comment sections on articles posted by the Fox News Facebook page frequently attract individuals that agree with the attitudes championed by Fox News (conservative) – when a liberal user enters the comment section to disagree, particularly in an uncivil manner, conservative commenters react as if they are in a tightly-knit group that is experiencing deviance. An increasingly common reaction in such a scenario is to label the perceived deviant poster a “troll.” As such, it is important to consider how the “troll” label and the assumption of norm-deviance may overlap, and what reactions arise in the face of said norm-deviance.

As previously illustrated, group members react to perceived offline counter-attitudinal deviance in numerous ways. Social-psychological researchers have begun to investigate this phenomenon in an online context, as well. Much like their offline counterparts, online group members may respond to such deviance through various methods, such as ignoring/excluding the target, expressing disapproval of the target, or distancing themselves from the original group (Birchmeier, Joinson, & Dietz-Uhler, 2005; Jetten & Hornsey, 2014; Nicholls & Rice, 2017).

Ditrich and Sassenberg (2017) demonstrated responses to an opinion-based deviant with an online group paradigm. Participants in this research felt that the in-group deviant who expressed a counter-attitudinal opinion undermined the core of the group's identity, which made the group less relevant to the participants' own individual self-concepts. In response to this perceived identity subversion, participants supported literally excluding the target, or exclusion through derogation of the target as an attempt to re-affirm the group's homogeneity (Ditrich & Sassenberg, 2017).

Although the majority of deviance research regarding group processes has occurred offline thus far, the prevalence of group formation on SNS calls for increasing investigation into this phenomenon in an online context.

Integrating the Concept of “Trolling”

It is still unknown why people are motivated to use the “troll” label against online discussants who disagree, particularly when incivility is present. The previously outlined literature suggests a few things: (1) language and labels are powerful tools used to assign people into categories, and this categorization has consequences; (2) people frequently have defensive cognitions in the face of counter-attitudinal information, which is nearly unavoidable on SNS; (3) discussion incivility has a plethora of unwanted effects on both involved discussants and

third-party witnesses, and often reduces the possibility of reaching a deliberative consensus; and (4) groups that form due to shared similarity are frequent online, and while offline studies have illustrated responses to opinion-based deviance, research is beginning to demonstrate this same method of responding in an online context.

What does “trolling” have to do with this set of literature? Imagine you seek out a pro-attitudinal article on the issue of Brett Kavanaugh’s appointment to the Supreme Court on Facebook. Interested in what fellow pro-attitudinal individuals are saying about this issue, you enter the comment sections, expecting to read a discussion that you agree with. Instead, the very first comment is from a rather uncivil poster who vehemently disagrees with the article, with you, and with everyone else in the comment section. It is very likely that at least one individual will respond to this counter-attitudinal comment by accusing them of being a “troll.”

If comment sections function similarly to a group-setting for like-minded individuals, then announcing that an individual is a “troll” is a form of derogation that may function to prevent identity subversion. Additionally, categorizing the source of the uncivil, counter-attitudinal opinion as a “troll” may serve as a defense against the consequences of both incivility and counter-attitudinal exposure. Rather than thoughtfully engaging and deliberating with this counter-attitudinal discussant, it takes less effort to write them off as a “troll.” Furthermore, the effects of incivility and counter-attitudinal exposure may become dampened.

Current Research

The current research aimed to synthesize the findings of the aforementioned literature to analyze the effects of witnessing the “troll” label being applied to a counter-attitudinal discussant in a SNS group setting. The current research paradigm employed a pro-attitudinal online group setting intended to discuss the acceptability of transgender identities. Participants were

incidentally exposed to a single counter-attitudinal discussant within the online group. In order to investigate the impact and interaction of discussion civility and specific derogation label, a 2 (civil discussion vs. uncivil discussion) x 3 (“troll” label vs. “rude” label vs. no label) between-subjects design was used.

The manipulation of the “troll” or “rude” label was employed through other pro-attitudinal discussants rather than the participant themselves. This is to ensure that the manipulation is realistic and similar to witnessing the “troll” label being applied by another online user in a real comment section/discussion thread. It would be unrealistic to force participants to use a specific label themselves in this paradigm. A predominant assumption of this research is that adopting the conception that a counter-attitudinal commenter is a “troll” is strong enough to impact the outcomes of incivility, deviance, and counter-attitudinal exposure. Adopting this perception could occur through self-generated processes, or it could occur through a pro-attitudinal group member signifying that a counter-attitudinal poster belongs to the “troll” category. However, future research should aim to determine whether the effect of a participant using the “troll” label him/herself is different from the effect of witnessing a pro-attitudinal group member use it.

Research on the impact of uncivil discussion exposure suggests several possible psychological outcomes, including those that reduce the impact of counter-attitudinal information. Yet, the “troll” label may function to exacerbate these outcomes, and additionally make the counter-attitudinal information even less impactful. In other words, incivility research suggests that witnessing an uncivil, counter-attitudinal discussant may elicit negative emotions such as anger and anxiety, as well as a decrease in open-mindedness and an increase in attitude certainty (which bolster a defense against the counter-attitudinal information). However, if one

adopts the conception that this uncivil, counter-attitudinal discussant is simply a troll, then their opinion should not be taken seriously. Theoretically, this makes the opinion far less threatening. This should reduce negative affect, as well as intensify the decrease in open-mindedness and the increase in attitude certainty. In this sense, the counter-attitudinal information should have an even less deliberative impact on the witness, when he/she believes it is coming from a “troll.”

Furthermore, the “troll” label presumes that the target has specific intentions to be upsetting, to be funny, or use deception in order to “bait” fellow discussants into a fruitless argument. This intention assumption provides an explanation for why incivility may have been used. Therefore, this label may potentially have different effects than another ad-hominem label, such as “rude,” which does not inherently include assumptions about the target’s intentions. Said differently, someone who is “rude” is often a negative presence to the discussion, but they may still be a genuine actor who intends to successfully convince others of their true beliefs. These hypotheses inspire the following predictions:

H1a: Participants in the uncivil discussion condition will report higher ratings of anger and anxiety than those in the civil condition.

H1b: Within the uncivil condition, participants who witness the “troll” label will report lower ratings of anger and anxiety than those who witness the “rude” label or no label.

H2a: Participants in the uncivil discussion condition will report higher ratings of attitude certainty than those in the civil condition.

H2b: Within the uncivil condition, participants who witness the “troll” label will report higher ratings of attitude certainty than those who witness the “rude” label or no label.

H3a: Participants in the uncivil discussion condition will report lower ratings of open-mindedness than those in the civil condition.

H3b: Within the uncivil condition, participants who witness the “troll” label will report lower ratings of open-mindedness than those who witness the “rude” label or no label.

Research on effects of incivility exposure further suggests that such exposure results in higher intentions to participate, potentially because discussants find uncivil discussion to be more entertaining. I hypothesize that the “troll” label application will result in an additional increase in intention to participate. This could be for two reasons: (1) the presence of a troll may be entertaining to some discussants, or (2) properly “outing” a person who is a troll reduces their threat, increasing other participants’ desire to participate.

H4a: Participants in the uncivil discussion condition will report higher intentions to participate in the discussion than those in the civil condition.

H4b: Within the uncivil condition, participants who witness the “troll” label will report higher intentions to participate in the discussion than those who witness the “rude” label or no label.

While previous research has demonstrated that uncivil discussion exposure decreases ratings of deliberative potential (Hwang, Kim, & Huh, 2014), and the “troll” label may act as a defense to this outcome, it is also possible that the assumed presence of a “troll” may cause participants to expect lower deliberative potential overall. In other words, a “troll’s” goal may be to upset others and/or disrupt conversation, which inherently reduces the discussion’s deliberative potential. Although the use of the “troll”

label sufficiently identifies the “troll’s” existence, this does not prevent the supposed “troll” from continuing to disrupt the discussion.

H5a: Participants in the uncivil discussion condition will report lower ratings of perceptions of deliberative potential than those in the civil condition.

H5b: Within the uncivil condition, participants who witness the “troll” label will report lower ratings of deliberative potential than those who witness the “rude” label or no label.

Finally, group deviance research suggests that derogation works as a defense to a deviant’s presence. According to Eidelman and Biernat (2003), deviance derogation and group disidentification act as similar techniques in the face of group deviance. Therefore, participants who witness no deviance derogation through the application of either label should report less identification with the group. Additionally, a goal of this thesis is to determine if the labels “rude” and “troll” may differ in their impact on group dynamics. Potentially, identifying and labeling the deviant a “troll” would strengthen the group, as they have successfully outed someone whose intention is to disrupt and upset other group members. Thus, witnessing the “troll” label may result in even stronger group identification than witnessing the “rude” label.

H6a: Participants who do not witness the deviant being derogated will show less identification with the group than those who witness the deviant labeled a “troll” or “rude”.

H6b: Participants who witness the deviant labeled “rude” will show less identification with the group than those who witness the deviant labeled a “troll.”

Method

Participants

Four hundred and seventy-five undergraduate students at the University of Arkansas completed this study for credit in their General Psychology course ($M_{Age} = 19.3$; White/Caucasian = 80.4%, Black/African American = 6.9%, Other = 12.7%; Female = 64.8%, Male = 34.5%, Gender neutral/fluid = 0.4%; Supportive of transgender identities = 58.9%).

Procedure

Participants were brought to the lab and instructed to use a computer for the duration of the study. Each participant was first given a definition of being transgender that states: “Someone who is transgender identifies with a gender that they were not given at birth. For example, a transgender man is someone who was assigned ‘female’ at birth, but now identifies as male.” Then, participants were prompted to choose which of the following options they most closely agree with: “I am against a person identifying with a gender they were not given at birth (transgender)”, or “I am supportive of a person identifying with a gender they were not given at birth (transgender)”. Additionally, they indicated on a 4-point Likert-type scale how strongly they believe in this attitude (1 – *somewhat*, 4 – *extremely*). Participants were told that this information would help to place them in a discussion group that best matches their indicated point of view. Their response was then used to determine which online group they experienced for the remainder of the study. Their indicated attitude on transgender identities determined the comments expressed by the pro-attitudinal discussants, as well as the counter-attitudinal discussant in the following paradigm. Participants experienced a 90 second delay before being placed into the group, in order to strengthen the cover story that they are interacting with real people in real-time.

Participants were led to believe that they were joining a real online discussion group, with real members participating in the discussion in real-time. They were told that the purpose of this experiment was to “beta-test a new chat forum... meant to foster an environment to talk about thought-provoking topics with fellow University of Arkansas students”. However, the other group members in this study did not actually exist, and their responses were pre-determined and only varied by condition. All responses from these surreptitiously “real” discussants were all pro-attitudinal according to the participant’s indicated attitude, *except for one*. It was crucial that participants believed they were discussing this topic with fellow University of Arkansas students. This was to ensure that their in-group became salient, and that they were aware that the counter-attitudinal commenter (deviant) was a member of their in-group.

The online discussion was formatted to mirror a typical online discussion. Responses from other “members” appeared in 30 to 45 second intervals. All discussants including the participant were provided with a temporary, anonymous profile that corresponded with a color. This ensured that no identifying factors such as gender or race are revealed to participants. Anonymity was *not* manipulated, as it was not a goal to investigate the effect of anonymity. However, it was used within the paradigm to ensure that race and gender did not impact any of the outcome variables.

Participants were instructed to carefully read each comment. Additionally, they were told that they could give input to the discussion but had to wait their turn in order to foster a respectful discussion and encourage full comprehension of others’ responses. Participants could write a contribution to the discussion only once, after four pro-attitudinal messages occurred, and right before the counter-attitudinal message appeared. The exact wording and order of the

messages displayed to participants, including how they vary by condition, can be viewed in Tables 42-45.

After exposure to the discussion, participants were redirected to a set of questionnaires meant to investigate the dependent variables. Participants were first prompted with a modified, shortened version of the SPANE-N measure (Diener et al., 2009). Although this measure includes four emotions, the only target emotions that were later analyzed were anger and anxiety. As a manipulation check, participants then used a 7-point Likert scale (1 – *strongly disagree*; 7 – *strongly agree*) to indicate how “uncivil and impolite” they found the comments in their group discussion to be. Next, participants indicated their current level of attitude certainty and open-mindedness (Borah, 2014; observed $\alpha = 0.72$ & $\alpha = 0.44$ respectively), and how much deliberative potential they believed the discussion had (Hwang, Kim, & Huh, 2014; observed $\alpha = .70$). To investigate participants’ intention to continue participating in the same discussion, participants were told that they may randomly be given the chance to participate in the discussion with their matched discussion partners again in the future. Participants then used a 7-point Likert scale (1 – *strongly disagree*; 7 – *strongly agree*) to respond to the following statement: “I am interested in having future discussions with my matched discussion partners.” Then, participants responded to Luhtanen’s and Crocker’s Collective Self Esteem subscale and indicated their level of identification with the following three identities: a University of Arkansas student, a member of the particular discussion, and a person with their indicated opinion (supportive or not supportive of transgender identities) (1992; observed $\alpha = 0.81$).

Following the dependent variables questionnaires, participants reported their age, gender identity, ethnicity, religious affiliation, political affiliation, and political party affiliation. Finally, participants were asked if they noticed/suspected anything about the research, as an attempt to

exclude anyone who did not believe the cover story. After they indicated this, they were thoroughly debriefed on the deception used.

Results

For each of the dependent variables within this study, a primary analysis including all participants was conducted. Then, two exploratory analyses were conducted to investigate any moderating effects of participant gender and participants' pre-existing attitude towards transgender identities. Only significant interactions with the independent variables are discussed for these exploratory analyses. Finally, roughly 10% of the sample (49 out of a total of 475) was confident that they were talking with entirely fake participants, as opposed to other real discussants. Therefore, a final analysis was conducted for each dependent variable with all suspicious participants removed. Only results that differed from the full sample are discussed. The results for each dependent variable are as follows.

Manipulation Check

Firstly, a 2 X 3 between-subjects ANOVA indicated that participants perceived the uncivil discussion as being significantly more uncivil ($M = 4.30$, $SE = 0.09$) than participants in the civil discussion ($M = 2.92$, $SE = 0.09$), $F(1,469) = 113.70$, $p < 0.001$, $\eta^2 = 0.194$. There was no significant effect of label condition on perception of incivility, nor an interaction. Therefore, the civility manipulation was successful. This pattern of results held true even when suspicious participants were removed. See Tables 1 and 38 for all inferential statistics involving the civility manipulation check.

Anger

A primary 2 X 3 between-subjects ANOVA was conducted to investigate the effect of civility and label used on anger. This primary analysis indicated a significant effect of civility

condition, $F(1,469) = 10.25, p = 0.001, \eta^2 = 0.021$. Consistent with hypothesis H1a, participants in the uncivil condition reported significantly more anger ($M = 2.01, SE = 0.06$) than in the civil condition ($M = 1.73, SE = 0.06$). However, this analysis indicated no significant interaction between condition and label, which is inconsistent with hypothesis H1b. The removal of suspicious participants did not impact this effect. An exploratory 2 X 3 X 2 ANOVA indicated an interaction between pre-existing attitudes towards transgender identities and civility on anger, $F(1,460) = 8.56, p = 0.004, \eta^2 = 0.017$. Follow-up simple effects tests showed that within the uncivil condition, participants who were supportive of transgender identities reported significantly more anger ($M = 2.26, SE = 0.08$) than those who were non-supportive of transgender identities ($M = 1.63, SE = 0.10$). This difference was non-significant within the civil condition. See Tables 2, 17 – 19 and 28 for all inferential statistics involving anger.

Anxiety

Analyses investigating anxiety as a dependent variable indicated no significant effect of civility or label, nor an interaction, even with suspicious participants removed. This was inconsistent with both hypotheses H1a and H1b. See Tables 3 and 29 for all inferential statistics involving anxiety.

Attitude Certainty

A primary 2 X 3 between-subjects ANOVA was conducted to investigate the effect of civility and label used on attitude certainty. This primary analysis indicated a significant effect of civility condition, $F(1,469) = 4.53, p = 0.034, \eta^2 = 0.009$. Consistent with hypothesis H2a, participants in the uncivil condition reported higher attitude certainty ($M = 5.24, SE = 0.08$) than those in the civil condition ($M = 5.00, SE = 0.08$). Furthermore, this primary analysis indicated a significant effect of label condition, $F(2,469) = 3.70, p = 0.025, \eta^2 = 0.015$. A follow-up Tukey's

HSD post-hoc test was employed to test for differences in attitude certainty among label conditions. Participants in the rude label condition reported significantly lower attitude certainty ($M = 4.90$, $SE = 0.10$) than those in the no label condition ($M = 5.25$, $SE = 0.10$). Additionally, those in the rude label condition reported marginally significantly lower attitude certainty than those in the troll label condition ($M = 5.21$, $SE = 0.10$). However, this analysis indicated no significant interaction between condition and label, which is inconsistent with hypothesis H2b. The ANOVA accounting for the removal of suspicious participants indicated that significant effect of label condition on attitude certainty still held true, but the effect of civility condition did not, $F(1,417) = 2.27$, $p = 0.23$, $\eta^2 = 0.003$ (civil: $M = 5.09$, $SE = 0.09$; uncivil: $M = 5.24$, $SE = 0.08$). See Tables 4, 5, and 30 for all inferential statistics involving attitude certainty.

Open-Mindedness

A primary 2 X 3 between-subjects ANOVA showed no significant effect of civility or label, nor an interaction, on open-mindedness, which was inconsistent with both hypotheses H3a and H3b. However, the ANOVA accounting for the removal of suspicious participants revealed a significant effect of label condition, $F(2,417) = 3.03$, $p = 0.049$, $\eta^2 = 0.014$. A follow-up Tukey's HSD post-hoc test was employed to test for differences in open-mindedness among label conditions. Participants in the troll label condition reported significantly more open-mindedness ($M = 4.04$, $SE = 0.11$) than those in the no label condition ($M = 3.65$, $SE = 0.11$). Participants in the rude label condition ($M = 3.81$, $SE = 0.12$) did not significantly differ in open-mindedness when compared to the two other conditions. See Tables 6, 31, and 32 for all inferential statistics involving open-mindedness.

Intentions to Participate

A primary 2 X 3 between-subjects ANOVA was conducted to investigate the effect of civility and label used on participants' intentions to participate in the discussion group in the future. The primary analysis indicated a significant effect of civility condition, $F(1,469) = 5.48, p = 0.020, \eta^2 = 0.011$. Consistent with hypothesis H4a, participants in the uncivil condition reported higher levels of participation intention ($M = 5.18, SE = 0.10$) than those in the civil condition ($M = 4.84, SE = 0.10$). Additionally, the primary analysis indicated a marginally significant interaction between civility and label conditions, $F(2,469) = 2.96, p = 0.053, \eta^2 = 0.012$. A follow-up simple effects test indicated a significant difference in participation intention between the label conditions within the uncivil condition, but NOT within the civil condition. In other words, label condition had a differential effect for participants also in the uncivil condition, but not for those in the civil condition. A Tukey's HSD post-hoc test was employed to test for differences in intentions to participate among label conditions, specifically within the uncivil condition. The results indicated that participants in the uncivil condition who witnessed the troll label being used reported marginally higher intentions to participate ($M = 5.55, SE = 0.17$) than those in the rude ($M = 4.99, SE = 0.19$) and no label ($M = 5.00, SE = 0.18$) conditions. Although these results were marginally significant, they were consistent with hypothesis H4b. Further moderation analyses suggest that this interaction was somewhat driven by participants who are supportive of transgender identities. In other words, moderation analyses found a significant interaction between attitude towards transgender identities and label, $F(2,463) = 3.08, p = 0.047, \eta^2 = 0.012$. This effect showed a similar pattern to the primary analysis, such that the troll label boosted intention to participate for participants who are supportive of transgender identities ($M = 5.46, SE = 0.16$) when compared to those who were not supportive of transgender identities also

within the troll label condition ($M = 4.67$, $SE = 0.20$). Follow-up simple effects tests indicated that intention to participate only significantly differed between supportive and non-supportive participants in the troll label condition. A final analysis was conducted with suspicious participants removed. This subsequent analysis no longer indicated a significant effect of civility, nor a significant interaction. Although the marginal means changed ever so slightly when suspicious participants were removed, this could partially be due to a smaller sample size, and therefore less power. See Tables 12 – 15, 20 – 25, and 34 for all inferential statistics involving intention to participate.

Deliberative Potential

Analyses investigating participants' perception of the deliberative potential of the discussion as a dependent variable indicated no significant effect of civility or label, nor an interaction, even with suspicious participants removed. This was inconsistent with both hypotheses H5a and H5b. See Tables 7 and 33 for all inferential statistics involving perception of deliberative potential of the discussion.

University of Arkansas Student Identification

Analyses investigating the importance of being a University of Arkansas student to participants' identities as a dependent variable indicated no significant effect of civility or label, nor an interaction, even with suspicious participants removed. This was inconsistent with both hypotheses H6a and H6b, as they pertain to University of Arkansas student group membership. See Tables 8 and 35 for all inferential statistics involving University of Arkansas group membership.

Discussion Group Member Identification

A primary 2 X 3 between-subjects ANOVA was conducted to investigate the effect of civility and label used on the importance of being a member of the current discussion group to participants' identities. The primary analysis indicated a significant effect of civility condition, $F(1,469) = 11.73, p = 0.007, \eta^2 = 0.015$. Overall, participants in the uncivil condition reported that being a member of the discussion group was more important to their identity ($M = 3.13, SE = 0.08$) than those in the civil condition ($M = 2.90, SE = 0.08$). This effect held even when suspicious participants were removed from the analysis. No specific hypothesis was predicted about the effect of civility on group membership. Furthermore, this analysis did not indicate any significant effect of label condition, which is inconsistent with both hypotheses H6a and H6b, as they pertain to discussion group membership. See Tables 9 and 36 for all inferential statistics involving discussion group membership.

Pre-Existing Attitude Identity Importance

A primary 2 X 3 between-subjects ANOVA was conducted to investigate the effect of civility and label used on the importance of being a person with the specified pre-existing attitude towards transgender identities to participants' own identities. The primary analysis indicated a marginal effect of label condition, $F(2,469) = 2.53, p = 0.081, \eta^2 = 0.011$. A Tukey's HSD post-hoc test indicated that participants in the troll label condition felt that their pre-existing attitude was marginally more important to their identity ($M = 4.25, SE = 0.11$) than those in the no label condition ($M = 3.91, SE = 0.11$). No other comparisons were trending towards significant. Although these results were marginally significant, they were consistent with hypothesis H6a, as it pertains to attitudes towards transgender identities group membership. However, these results were inconsistent with hypothesis H6b, because the troll label and rude

label conditions did not significantly differ. This marginal effect of label condition appeared to be related to gender during this study. A moderation analysis revealed a significant interaction between participant gender and label condition, $F(2,460) = 3.38, p = 0.035, \eta^2 = 0.014$. Follow-up simple effects tests indicated that, within the no label and the rude label conditions, women felt that their pre-existing attitude towards transgender identities was significantly more important to their identities than men. On the other hand, men and women reported similar levels of attitude importance in the troll label condition. It appears that the troll label had differential effects on men in the sample than for women. It is important to note that the marginal effect of label condition found in the primary analysis became nonsignificant when removing suspicious participants, $F(2,417) = 1.59, p = 0.20$. See Tables 10, 11, 26, 27, and 37 for all inferential statistics involving attitude towards transgender identities group membership.

Pre-Existing Attitude Strength

Finally, an independent samples t-test was conducted to investigate the difference in attitude strength between participants who were supportive and not supportive of transgender identities. Attitude strength was indeed significantly different among participants, $t(473) = -2.107, p = 0.036$. It is important to note that sample sizes for this analysis were unequal (supportive: $N = 280$; non-supportive: $N = 195$). However, supportive participants felt significantly more strongly about their stance on transgender identities ($M = 2.97, SE = 0.06$) than non-supportive participants ($M = 2.77, SE = 0.72$). Implications of this difference and its possible impact on moderating analyses are discussed subsequently. See Table 16 for inferential statistics involving difference in attitude strength between supportive and not-supportive participants.

Discussion

Brief Restatement of Results

In the primary analysis, the two independent variables in this study had significant effects on several hypothesized dependent variables, as well some dependent variables that were not previously anticipated. Overall, the presence of incivility increased participants' feelings of anger and attitude certainty, intentions to participate in the discussion again, and the importance of being a discussion member to one's identity. This supported hypotheses H1a (for anger only), H2a, and H4a. The rude label decreased participants' attitude certainty regardless of incivility condition. The troll label, on the other hand, marginally increased how important participants felt that their pre-existing attitude about transgender individuals was to their own identity. This partially supported hypothesis H6a. This was also regardless of incivility condition.

The predicted interaction between the troll label and incivility only occurred for intentions to participate, and this effect was marginal. This partially supported hypothesis H4b. The elicited pattern suggests that participants desired to participate the most when incivility was present, and the troll label was used. No other interactions between these two variables were significant.

Moderating analyses were used investigate the role of gender, as well as participants' pre-existing attitudes towards transgender identities. Gender moderated the effect of the troll label on the importance of the pre-existing attitude to one's own identity, such that male participants assigned significantly more importance than other participants in this condition. Additionally, participants who were supportive of transgender identities were significantly angrier in the face of incivility, and were more willing to participate in the near future when the troll label was used.

These effects have implications for the dynamics underlying the topic used in this study (transgender identities) and how the troll label is perceived.

Finally, suspicious participants were removed in a subsequent analysis. Incivility no longer had a significant effect on attitude certainty, nor intention to participate. Additionally, the marginal interaction between label and civility on participation was no longer significant. However, this analysis revealed a significant effect of label on open-mindedness, such the troll label significantly increased open-mindedness as compared to the no label condition.

See Table 41 for all a priori hypotheses, and whether they were supported or not supported in the primary analysis. All aforementioned results are unpacked in the following discussion.

Main Effects of Incivility

Extent research on computer-mediated communication has demonstrated far-reaching effects of the presence of incivility (Anderson et al., 2014; Borah, 2014; Gervais, 2017; Graf, Erba, & Harn, 2017; Hwang, Kim, & Huh, 2014; Hwang, Kim, & Kim, 2018; Kim, 2015; Weber, Prochazka, & Schweiger, 2017). This current study found similar effects for some dependent variables, failed to find similar effects for others, and found surprising effects that were not previously anticipated.

Previous research has demonstrated that uncivil discussion in a computer-mediated context can elicit negative affect, including anger and anxiety (Gervais, 2015; Gervais, 2017). This finding was partially replicated in this study. Participants in the uncivil condition reported significantly more anger, but neither condition elicited anxiety. In fact, participants in this study reported generally low anger overall, and virtually no anxiety. This could potentially be a consequence of participants not feeling immersed or invested enough during the study. On the

other hand, the topic of transgender identities may have been enough to induce anger, but not necessarily induce anxiety, for this cisgender sample.

Attitude certainty and open-mindedness are often seen as two sides of the same coin – however, they did not quite function that way in this specific study. As hypothesized, exposure to discussion incivility increased attitude certainty in the primary analysis. Although people may consider online discussions as opportunities to convince or persuade the other side, results of this study and previous studies show evidence of the opposite effect when incivility is present. Open-mindedness was hypothesized to function inversely to attitude certainty. However, there were no initial significant effects of open-mindedness during the primary analysis. Therefore, while discussion incivility resulted in heightened attitude certainty, it did not concurrently result in decrements in open-mindedness.

As hypothesized, participants in the uncivil condition reported significantly higher intentions to participate in the same discussion group in the near future. This appears to be like a “train wreck” effect – it’s uncivil and somewhat unpleasant, but it’s entertaining to watch! This could partially explain why discussants are still attracted to contentious and uncivil debates on social media, despite increases in negative affect and the frequent lack of consensus reached. On the other hand, the presence of fellow pro-attitudinal discussion members may have helped participants feel comfortable participating, especially in the presence of incivility. Perhaps if the discussion was one-on-one, and the participant was required to defend against the counter-attitudinal discussant alone, they would feel less desire to participate again.

Surprisingly, incivility did not impact participants’ perceptions of the discussion’s deliberative potential. Theoretically, the presence of an uncivil discussant should decrease the ability to reach a deliberative consensus during a discussion (Hwang, Kim, & Huh, 2014;

Hwang, Kim, & Kim, 2018). There are a few possible reasons why this was not the case in this study. Perhaps participants felt that the overall format and length of the discussion was not conducive to having a full, thoughtful conversation. In fact, a portion of participants reported that they felt upset that the conversation ended so quickly, and that they did not get a chance to write a second comment. Being cut off in the middle of the conversation may have decreased participants' perception of the discussion's deliberative potential regardless of condition.

Alternatively, many heated discussions involve a handful of discussants on both sides of the issue. In this study, there was only one counter-attitudinal discussant in the conversation. For incivility to have a significant impact on perceptions of deliberative potential, perhaps there needs to be multiple uncivil counter-attitudinal discussants present. Finally, this lack of effect may also be related to the topic at hand. Perhaps participants did not feel that the hyper-polarized topic of transgender identities even *has* a consensus to be reached!

Intriguingly, participants in the uncivil condition overall felt that being a member of this discussion was *more* important to their identity. While this was not anticipated by any hypothesis, it deserves further consideration. It is possible that the presence of an uncivil counter-attitudinal discussant made the participant feel closer with the rest of the group, regardless of any derogation. In other words, the discourteous member may have influenced the rest of the group to "unite" against their presence. This effect was anticipated, but only in conjunction with a derogation label being used.

Since there was no significant difference in the uncivil condition between the label conditions, this suggests that this increase in identification with the discussion group occurred without derogation being necessary. In this study, pro-attitudinal discussants responded to the uncivil counter-attitudinal discussant to try to "correct" their incorrect point of view. It is

possible that witnessing the group “correct” the deviant member when they were being uncivil motivated participants to subsequently identify with the discussion group.

Main Effects of Label

A priori hypotheses about main effects of label usage only pertained to group identification variables. However, label usage had a main effect on attitude certainty, which was not anticipated prior to conducting this research.

Overall, the rude label lessened attitude certainty significantly, when compared to the troll label condition and the no label condition. Importantly, this indicates that the labels “rude” and “troll” function differently – participants are psychologically responding to them in a distinct manner. However, the rude label having a stronger impact on attitude certainty is somewhat antithetical to the original hypotheses. The connotations behind being rude and trolling are rather different. Rude discussants are impolite, but often are genuine. Trolls on the other hand are impolite, but also deceptive and sadistic (Buckels, Trapnell, & Paulhaus, 2014; Shachaf & Hara, 2010). Nonetheless, participants let their guards down significantly more when the disagreeing discussant was labelled “rude.” Perhaps when participants witnessed the troll label being used, they became more vigilant and cautious because the label convinced them the disagreeing discussant would try to deceive and upset them on purpose. This process could be reflected by higher attitude certainty ratings in the troll condition. If this is the case, then troll label usage is not functioning as a defense against the increase in attitude certainty. Rather, it is enhancing such feelings, because participants feel guarded and defensive around someone who is labelled a troll.

Regarding group identification, label conditions had no significant impact on participants’ identification with being a University of Arkansas student, nor with being a member of the discussion. Although it was hypothesized that the lack of derogation against the deviant

member would lead to less identification with the group, this was not the case for this specific group membership. Although it was made explicit that all discussants were fellow university students, it is possible that participants did not take note of this. Therefore, this group membership may not have been salient. Alternatively, these specific labels simply may not have a significant impact on this specific form of group identification.

Unlike University of Arkansas group membership, discussion group membership should have been very salient. Perhaps participants needed to be a member of this discussion group for longer periods of time for label derogation to have a differential effect on discussion group identification. Additionally, the counter-attitudinal discussant did not attack the participants' university group membership, nor discussion group membership. These types of group memberships were not directly threatened in this study. Therefore, if participants did not feel their group membership was threatened on these dimensions in the first place, then derogation may not have been necessary in order to reaffirm group membership identification.

This study also investigated how much being a person with a certain pre-existing attitude about transgender identities is important to a participant's own identity. This form of group membership is more implicit – people often do not explicitly choose to join this group, and yet they form tight bonds and demonstrate intergroup conflict on the basis of shared attitudes. This is especially the case on social media comment sections that discuss contentious topics. For this implicit group membership, the troll label marginally increased participants' identification when compared to the no label condition, while the rude label did not.

This label impacted only this form of group membership identification, and not the other two. In a way, this makes sense – the content of the counter-attitudinal discussant's comment was directly about the pre-existing attitude, and *not* about being a university member, nor a

discussion group member. This is also the case when the “troll” label is used ecologically on social media. People tend to apply this label to those who disagree with them on a pre-existing attitude, such as transgender identities, abortion, marijuana, religion, presidential candidates, etc. Therefore, using this specific derogatory label in this circumstance may have a more direct impact on this implicit group membership, because the content of the deviant comment is unambiguously related to these pre-existing attitudes. In other words, when someone’s pre-existing attitude is under attack, and they see that person labeled a “troll,” they may once again feel more comfortable identifying with that pre-existing attitude group. But this may not affect discussion group membership nor university membership as strongly, because these specific groups were not directly under attack by the counter-attitudinal discussant in this study.

Interactions & Moderating Effects

Results provided evidence for a (marginal) interaction between label and civility for only one hypothesized dependent variable. Moderating analyses were employed to further explore how gender and pre-existing attitudes towards transgender identities may dynamically impact the manifestation of the investigated variables.

Firstly, although incivility induced anger, a lack of significant interaction in the preliminary analysis suggests that neither label was sufficient in reducing this anger. Therefore, seeing someone labeled as “rude” or a “troll” may not be enough to quell the negative emotional aftermath of counter-attitudinal incivility.

The significant moderating interaction between participants’ pre-existing attitudes towards transgender identities and condition suggests that participants who are supportive of transgender identities were angered far more by incivility from a counter-attitudinal discussant than participants who are non-supportive of transgender identities. This is additionally

complicated by the fact that supportive participants felt significantly more strongly about their attitude towards transgender identities overall.

These results seem to suggest that anger is more likely to be elicited from an uncivil discussion when someone feels more strongly about the issue being discussed. In other words, if non-supportive participants do not feel very strongly about their attitude, a disagreeing discussant is not as emotionally threatening. This further supports the notion in the online incivility research that uncivil discussions may be more psychologically impactful when they involve issues that are personally relevant and important to participants (Wang & Silva, 2017). Furthermore, although many topics are incredibly polarized in the current political climate, participants on one side of the issue may not feel as strongly as participants on the other side for every topic.

In the uncivil condition, participants were marginally more willing to participate in the discussion when the counter-attitudinal discussant was labeled as a “troll” than when they were labeled “rude,” or no label was used. Although marginal, this interaction effect was consistent with the a priori hypothesis. This result suggests that the troll label in combination with the presence of incivility slightly increases desires to participate in the “Hog Chat” discussion again. There are two possible reasons for this effect. Firstly, participants may have been unsure about participating in the future because the uncivil counter-attitudinal discussant was unpleasant and threatening. But when that discussant was successfully “outed” as a troll, participants felt more comfortable continuing their participation. Alternatively, this effect could be looked at it from the aforementioned “train wreck” perspective. Incivility overall could be entertaining for many participants; when you add a “troll” into the mix, the discussion becomes even more entertaining! Rather than the label acting as a defense against withdrawal from the discussion,

participants possibly were more interested in participating in a discussion that featured a troll-like member, because trolls like to cause trouble for fun.

Further moderation analyses suggest that this effect was partially driven by participants who were supportive of transgender identities. This portion of the sample wanted to participate significantly more after seeing the “troll” labeled used, but non-supportive participants did not change in participation intention. This is likely the case because supportive participants were more emotionally threatened during this study, and also reported that their pre-existing attitude was stronger overall. This further suggests that the effect of incivility and label usage may be heightened when participants feel strongly about the topic at hand.

Additionally, recall that participants reported that being a person with their pre-existing attitude towards transgender identities was marginally more important to their identity in the troll label condition than the no label condition. Moderation analyses suggested that the responses of male participants in this sample may have driven this marginal effect. This poses the question: do men react differently to the concept of “trolling” than other participants? Even today in 2020, men tend to dominate the spaces of the internet in which the word “trolling” is used more often, including massively multiplayer online (MMO) video gaming platforms and forums such as Reddit (Gough, 2018; Statista Research Department, 2016). Therefore, men likely have more experience with this term, and therefore a more informed understanding of the connotation behind this particular label. Consequently, male participants may have been more sensitive to the troll label as it appeared in this current study.

Removing Suspicious Participants

It was important during this discussion study to maintain as much psychological realism and believability as possible. The overarching goal of this study was to analyze participants’

reactions to participating in a real discussion. This differed from many previous studies in the literature that analyzed participants' reactions to *reading* online discussion comments (Anderson et al., 2014; Borah, 2014; Graf, Erba, & Harn, 2017; Hwang, Kim, & Kim, 2018; Wang & Silva, 2018). The first-person nature of discussion participation likely elicits different effects than the third-person nature of reading comments. Therefore, it was essential to craft a believable discussion while still maintaining control over the investigated variables. Furthermore, the funnel-debriefing method was critical, as it allowed a secondary analysis to determine if any of the previously stated effects (or lack thereof) changed when suspicious participants were removed.

The additional analysis with suspicious participants removed yielded different results for attitude certainty and open-mindedness. Firstly, the effect of civility on attitude certainty was no longer significant. This is potentially related to lost power, as removing a portion of the sample reduces the ability to detect a significant effect (McBride, 2020). Yet, mean attitude certainty decreased in both conditions upon removal of suspicious participants, with a *slightly* larger decrease in the uncivil condition. Therefore, as opposed to lost power as an explanation, it is possible that skepticism towards the authenticity of the study overall inflated attitude certainty in the primary analysis. This skepticism may have been exacerbated by the presence of incivility. In other words, participants may have become defensive and more certain of their pre-existing attitude as a response to feeling suspicious towards the research, in combination with civility condition. However, it is difficult to determine which component was responsible for this curious pattern of effects.

The preliminary analysis suggested no significant effect of either independent variable on open-mindedness. However, removal of suspicious participants complicated the picture. With

these participants removed, the troll label significantly increased open-mindedness overall, regardless of the presence of incivility. For this variable, the troll and rude labels again seem to present a different psychological function – but this time, the troll label had the significant impact on the variable in question, and the rude label did not. It appears that the troll label sufficiently increased open-mindedness, and the rude label sufficiently decreased attitude certainty. Not only is this evidence that these two labels function differently, but that attitude certainty and open-mindedness are distinct psychological processes as well.

Like the pattern of attitude certainty, it is possible that suspicious participants were feeling skeptical and defensive throughout the study. Therefore, they may have had low open-mindedness overall, regardless of condition. These participants may have been superficially lowering the troll label condition mean. Yet, there was still no significant interaction found between incivility and label. It was expected that incivility presence would be necessary to see increases in defensive variables, such as high attitude certainty and low open-mindedness. However, it appears that even civil disagreement is enough for participants to react differentially to label usage. In other words, even civil disagreement may be enough to cause discomfort, but incivility still likely exacerbates this discomfort.

When suspicious participants were removed, the interaction between label and civility on participation intention was no longer marginally significant. The marginal means of each label condition within the uncivil condition with suspicious participants removed do not differ greatly from the marginal means from the primary analysis. Therefore, it is possible that removal of 10% of the sample resulted in a loss of power. Since the original effect was marginal, this deduction in the analytic sample may have had an important impact. On the other hand, it is possible that suspicious people were the ones driving this effect. Since these participants were already aware

that they were talking to a computer, they may have found it additionally entertaining that the word “troll” was used. However, it’s a bit inexplicable that suspicious participants would *want* to continue participating in the near future. If they knew that all the fellow participants were fake, why would they want to participate again?

Final Take-Aways

The results of this study suggest several things about the investigated variables and other intertwining dynamics. Primarily, incivility had an impact on anger, attitude certainty, intentions to participate, and importance of being a discussion group member to one’s identity.

Additionally, the troll label seemed to increase open-mindedness and the importance of one’s opinion on transgender identities to one’s identity, while the rude label decreased attitude certainty. The troll label and incivility only interacted to have a marginal effect on intentions to participate.

Overall, it appears that the effects of civility and label usage may be dependent on the topic at hand. When participants care more about the topic, they may be more impacted by the presence of incivility, and/or which derogatory label is used.

For most variables, there was no significant interaction effect. This might suggest that simple disagreement, even when civil in nature, may be enough to illicit discomfort in participants. In turn, label usage may still come in handy when defending against this discomfort. There were multiple instances in which the “rude” label and the “troll” label had differential psychological effects, which strongly suggests that they do not function identically. Additionally, label usage appeared to directly impact only pre-existing attitude group membership in this study. This may be related to the content of the counter-attitudinal discussant’s comment.

Finally, removing suspicious participants allowed a more nuanced look into the effects at play. For attitude certainty, open-mindedness, and intentions to participate, it appears that suspicious participants were reacting to the study in a different way from non-suspicious participants. Therefore, it is important to utilize a form of funnel debriefing at the end of any deception study to better understand results.

Limitations

This study has a handful of limitations to its design. Firstly, with any research utilizing deception, there is the risk that participants will be cognizant of the deception used. While the fake comments from fictitious discussants were based on real-life comments found online, and while the computer interface during the study was crafted to appear as real as possible, roughly 10% of the sample was aware that they were not discussing with real people. Additionally, it is possible that more participants were suspicious of the authenticity of this study, but did not admit it during the funnel debriefing, and therefore were never removed during the subsequent analysis. Although very few participants self-reported that their suspicions impacted their responses, they may not be fully aware of how these suspicions may have impacted their behavior. Therefore, it is important to consider the ecological validity of the stated results when discussing them through the lens of genuine social media use. While the design was intended to be as realistic as possible to mirror everyday computer-mediated discussions, it is likely that participants did not feel as psychologically immersed throughout the study as was intended.

Secondly, the goal of this design was to maintain realism, but also manipulate the type of label used against the counter-attitudinal discussant. These two conflicting goals made it virtually impossible to manipulate this variable in a way that resulted in the participant themselves using the label. Therefore, it was manipulated such that the label came from a fellow

(fictitious) pro-attitudinal discussant. Conclusions about any significant effects of label condition or lack thereof can only provide insight into the psychological impacts of observing another pro-attitudinal discussant use the label. In other words, conclusions cannot be drawn about what psychologically occurs when the participant *themselves* use a specific label. This may speak to why there were no significant differences between the rude label and the troll label for some of the dependent variables, such as anger. Perhaps these two labels would impact the dependent variables discriminately if the participant personally employed them against the counter-attitudinal discussant. On the other hand, it is nonetheless possible that the function of the troll label is virtually indistinguishable from the function of the rude label for these dependent variables, despite the differing connotations behind both labels.

Thirdly, since this study was an amalgamation of different theories and branches of social psychology, it cannot act as a direct replication of any previous findings. Rather, it acts as a supplement to previous studies on incivility and group dynamics. Some hypotheses inspired by previous research were supported in this study, and some were not. This may possibly be due to the change in context and discussion content from original studies.

Previous online discussion incivility research has not investigated attitudes on transgender identities. This discussion topic may impact the examined variables slightly differently than previously investigated topics. Many previously studied topics like marijuana legalization, abortion, and gun rights are very personal to a large percentage of discussants. Most people understand these topics well, and even have personal experience associated with these topics. However, while the acceptance of transgender people is still a polarized subject, most people don't have personal experience with transgender people and their daily struggles. For

participants of this current study, discussing transgender identities was likely a cognitively abstract exercise.

Furthermore, group identification variables may have functioned slightly differently in this study because the group was online as opposed to in-person. Additionally, perhaps the duration of the group was not long enough for participants to feel immersed, included, and interested. This discrepancy between previous findings and current findings highlights the importance of context and discussion content in this line of research.

During the funnel-debriefing, a sizeable portion of participants suspected that the purpose of this study was to collect data on University of Arkansas students' attitudes towards transgender identities. In today's political climate, many citizens are concerned with the ulterior motives of so-called "PC culture." Many participants may have suspected that the results of this study would be used to implement policies they disagree with. For these participants, this mistrust could have impacted how forthcoming they were during the entirety of the study.

Finally, participants in this study were not asked how they personally define the concept of "trolling." This concept is still relatively new, and not everyone has a clear understanding of what it means. Although this study used a youthful, college-aged sample, it is still possible that a portion of the sample in the troll label condition did not know what the label means. For these participants, witnessing the label would not conjure cognitions about deceptive malintent from the counter-attitudinal discussant. In turn, this would directly impact how they would respond to the subsequent dependent variables. Differences in variables between the rude label and troll label conditions may have been more distinct if the entire sample was informed of the concept of "trolling."

Future Directions

As the first of its kind, this study was an attempt to merge several divisions of social psychology to examine a relatively new research question. Subsequently, there are several directions that future research can explore to continue this investigation.

Future research ideally should manipulate derogatory label usage so that it comes from the participant themselves. This would clarify whether personal label usage – as opposed to witnessing it from a fellow pro-attitudinal discussant – has a differential effect on the investigated variables. If the “troll” label is examined in future research, researchers should ensure that they are studying a sample that is clear and knowledgeable about what this label means. This could be done by using a pre-screening technique, or by including an open-ended question at the very end of the study.

Additionally, the more realistic the study is, the less suspicious participants are. Future research should attempt to craft an interface that mirrors social media as much as possible, feels real to participants, but still maintains proper control over extraneous variables. This would also increase participants’ feelings of immersion. Tangentially, including an in-person control condition would provide insight into the comparison between in-person discussion dynamics and online discussion dynamics.

Lastly, this current study examined a somewhat unique topic – transgender people. Results clearly demonstrated that participants were relatively evenly split on which side of this topic they were on. However, supportive participants’ attitudes were far stronger than non-supportive participants. This complicated the conclusions drawn about some of the study’s analyses. In the future, it would be beneficial to replace this topic with one that is more balanced. Researchers could do a preliminary study or use a pre-screening technique to determine how

their sample feels about their selected topic. They should determine that the topic has an even split, and both sides feel as strongly about their attitude.

In conclusion, this research adds support to the growing literature that demonstrates a strong effect of discussion incivility on anger, attitude certainty, and intentions to continue participating. Additionally, it appears that witnessing a counter-attitudinal discussant specifically labelled as a “troll” may increase a person’s open-mindedness and their intentions continue discussion participation, as well as motivate them to identify more strongly with their pre-existing attitude. This label also appeared to have a slightly different psychological impact than the “rude” label, which distinctly reduced participants’ attitude certainty. Furthermore, label usage impacted these variables regardless of an interaction with incivility. This suggests that a person may benefit from label usage even when the counter-attitudinal information is presented in a civil manner. This pattern of results may help researchers explain why Facebook users are often readily willing to adopt the perception that a disagreeing discussant is “trolling,” rather than attempting to have a deliberative and meaningful discussion.

References

- Anderson, A. A., Brossard, D., Scheufele, D. A., Xenos, M. A., & Ladwig, P. (2013). The “Nasty Effect”: Online incivility and risk perceptions of emerging technologies. *Journal of Computer-Mediated Communication*, *19*(3). doi:10.1111/jcc4.12009
- Becker, H. S. (1963). *Outsiders: Studies in the sociology of deviance*. Simon & Schuster.
- Birchmeier, Z., Joinson, A. N., & Dietz-Uhler, B. (2005). Storming and forming a normative response to a deception revealed online. *Social Science Computer Review*, *23*(1). doi:10.1177/0894439304271542
- Borah, P. (2013). Interactions of news frames and incivility in the political blogosphere: Examining perceptual outcomes. *Political Communication*, *30*(3), 456-473. doi:10.1080/10584609.2012.737426
- Brundidge, J. (2010). Encountering “difference” in the contemporary public sphere: The contribution of the internet to the heterogeneity of political discussion networks. *Journal of Communication*, *60*(4). doi:10.1111/j.1460-2466.2010.01509.x
- Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. John Wiley and Sons.
- Buckels, E. E., Trapnell, P. D., & Paulhus, D. L. (2014). Trolls just want to have fun. *Personality and Individual Differences*, *67*, 97-102. doi:10.1016/j.paid.2014.01.016
- Castano, E., Yzerbyt, V. Y., Bourguignon, D., & Seron, E. (2002). Who may enter? The impact of in-group identification on in-group/out-group categorization. *Journal of Experimental Social Psychology*, *38*(3), 315-322. doi:10.1006/jesp.2001.1512
- Caza, B. B., & Cortina, L. M. (2007). From insult to injury: Explaining the impact of incivility. *Basic and Applied Social Psychology*, *29*(4), 335-350. doi:10.1080/01973530701665108
- Chen, J. M., & Ratliff, K. A. (2015). Implicit attitude generalization from Black to Black-White biracial group members. *Social Psychological and Personality Science*, *6*(5), 544-550. doi:10.1177/1948550614567686
- Chen, J. M., Pauker, K., Gaither, S. E., Hamilton, D. L., & Sherman, J. W. (2018). Black + white = not white: A minority bias in categorizations of Black-White multiracials. *Journal of Experimental Social Psychology*, *78*, 43-54. doi:10.1016/j.jesp.2018.05.002
- Cho, D., & Kwon, K. H. (2015). The impacts of identity verification and disclosure of social cues on flaming in online user comments. *Computers in Human Behavior*, *51*, 363-372. doi:10.1016/j.chb.2015.04.046

- Coe, K., Kenski, K., & Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication, 64*(4), 658-679. doi:10.1111/jcom.12104
- Coles, B. A., & West, M. (2016). Trolling the trolls: Online forum users constructions of the nature and properties of trolling. *Computers in Human Behavior, 60*, 233-244. doi:10.1016/j.chb.2016.02.070
- Cooper, J., & Fazio, R. H. (1984). A new look at dissonance theory. *Advances in Experimental Social Psychology, 17*, 229-266. doi:10.1016/S0065-2601(08)60121-5
- Cortina, L. M., Magley, V., Williams, J. H., & Langhout, R. D. (2001). Incivility in the workplace: Incidence and impact. *Journal of Occupational Health Psychology, 6*(1), 64-80. doi:10.1037/1076-8998.6.1.64
- Craker, N., & March, E. (2016). The dark side of Facebook®: The Dark Tetrad, negative social potency, and trolling behaviours. *Personality and Individual Differences, 102*, 79-84. doi:10.1016/j.paid.2016.06.043
- De Clercq, D., Haq, I. U., Azeem, M. U., & Raja, U. (2018). Family incivility, emotional exhaustion at work, and being a good soldier: The buffering roles of waypower and willpower. *Journal of Business Research, 89*, 27-36. doi:10.1016/j.jbusres.2018.04.002
- Denegri-Knott, J., & Taylor, J. (2005). The Labeling Game: a conceptual exploration of deviance on the internet. *Social Science Computer Review, 23*(1). doi:10.1177/0894439304271541
- Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D., Oishi, S., & Biswas-Diener, R. (2009). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research, 97*(2), 144-156. doi:10.1007/s11205-009-9493-y
- Ditrich, L., & Sassenberg, K. (2017). Kicking out the trolls: Antecedents of social exclusion intentions in Facebook groups. *Computers in Human Behavior, 75*(C), 32-41. doi:10.1016/j.chb.2017.04.049
- Dylko, I., Dolgov, W. H., Eckhart, N., Molina, M., & Aaziz, O. (2018). Impact of customizability technology on political polarization. *Journal of Information Technology and Politics, 15*(1). doi:10.1080/19331681.2017.1354243
- Edwards, K., & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology, 71*(1), 5-24. doi:10.1037/0022-3514.71.1.5
- Eidelman, S., & Biernat, M. (2003). Derogating black sheep: Individual or group protection? *Journal of Experimental Social Psychology, 39*(6), 602-609. doi:10.1016/S0022-1031(03)00042-8

- Eidelman, S., Silvia, P. J., & Biernat, M. (2006). Responding to deviance: target exclusion and differential devaluation. *Personality and Social Psychology Bulletin*, 32(9), 1153-1164. doi:10.1177/0146167206288720
- Fasoli, F., Paladino, M. P., Carnaghi, A., Jetten, J., Bastian, B., & Bain, P. G. (2015). Not “just words”: Exposure to homophobic epithets leads to dehumanizing and physical distancing from gay men. *European Journal of Social Psychology*, 46(2). doi:10.1002/ejsp.2148
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.
- Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *The Journal of Abnormal and Social Psychology*, 58(2), 203. doi:10.1037/h0041593
- Festinger, L., Schachter, S., & Back, K. (1950). *Social pressures in informal groups: A study of human factors in housing*. Harper.
- Fichman, P., & Sanfilippo, M. R. (2015). The bad boys and girls of cyberspace: How gender and context impact perception of and reaction to trolling. *Social Science Computer Review*, 33(2), 163-180. doi:10.1177/0894439314533169
- Foster, G. G., Schmidt, C. R., & Sabatino, D. (1976). Teacher expectancies and the label “Learning Disabilities”. *Journal of Learning Disabilities*, 9(2). doi:10.1177/002221947600900209
- Freeman, S., & Algozzine, B. (1980). Social acceptability as a function of labels and assigned attributes. *American Journal of Mental Deficiency*, 84(6), 589-595.
- Garrett, R. K., Carnahan, D., & Lynch, E. K. (2013). A turn toward avoidance? Selective exposure to online political information, 2004-2008. *Political Behavior*, 35(1), 113-134. doi:10.1007/s11109-011-9185-6
- Garrett, R. K., & Stroud, N. J. (2014). Partisan paths to exposure diversity: Differences in pro- and counterattitudinal news consumption. *Journal of Communication*, 64, 680-701. doi:10.1111/jcom.12105
- Gelman, S. A., Collman, P., & Maccoby, E. E. (1986). Inferring properties from categories versus inferring categories from properties: The case of gender. *Child Development*, 57(2), 396-404. doi:10.2307/1130595
- Gervais, B. T. (2015). Incivility online: Affective and behavioral reactions to uncivil political posts in a web-based experiment. *Journal of Information Technology & Politics*, 12(2), 167-185. doi:10.1080/19331681.2014.997416
- Gervais, B. T. (2016). More than mimicry? The role of anger in uncivil reactions to elite political incivility. *International Journal of Public Opinion Research*, 29(3), 384-405. doi:10.1093/ijpor/edw010

- Gough, C. (2018). MOBA/MMORPG gamer share worldwide Q1 2018, by gender. *Statista*. Retrieved from <https://www.statista.com/statistics/871031/moba-mmorpg-gamers-by-gender/>
- Graf, J., Erba, J., & Harn, R. (2017). The role of civility and anonymity on perceptions of online comments. *Mass Communication and Society, 20*(4), 526-549. doi:10.1080/15205436.2016.1274763
- Hardaker, C. (2010). Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions. *Journal of Politeness Research, 6*(2), 215-242. doi:10.1515/JPLR.2010.011
- Helbig, K. (2014, April 20). 11 numbers that show how prolific illegal downloading is right now. *GlobalPost: Business, Economics and Jobs*. Retrieved from <https://www.pri.org/stories/2014-04-20/11-numbers-show-how-prolific-illegal-downloading-right-now>
- Herring, S., Job-Sluder, K., Scheckler, R., & Barab, S. (2002). Searching for safety online: Managing “trolling” in a feminist forum. *The Information Society, 18*, 371-384. doi:10.1080/01972240290108186
- Hogg, M. A., & Reid, S. A. Social identity, self-categorization, and the communication of group norms. *Communication Theory, 16*(1). doi:10.1111/j.1468-2885.2006.00003.x
- Hwang, H., Kim, Y., & Huh, C. U. (2014). Seeing is believing: Effects of uncivil online debate on political polarization and expectations of deliberation. *Journal of Broadcasting and Electronic Media, 58*(4), 621-633. doi:10.1080/08838151.2014.966365
- Hwang, H., Kim, Y., & Kim, Y. (2016). Influence of discussion incivility on deliberation: An examination of the mediating role of moral indignation. *Communication Research, 45*(2). doi:10.1177/0093650215616861
- Hwang, H., Pan, Z., & Sun, Y. (2008). Influence of hostile media perception on willingness to engage in discursive activities: An examination of mediating role of media indignation. *Media Psychology, 11*(1), 76-97. doi:10.1080/15213260701813454
- Ito, T. A., & Urland, G. R. (2003). Race and gender on the brain: Electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology, 85*(4), 616-626. doi:10.1037/0022-3514.85.4.616
- Jetten, J., & Hornsey, M. J. (2014). Deviance and dissent in groups. *Annual Review of Psychology, 65*, 461-485. doi:10.1146/annurev-psych-010213-115151
- Kathurwar, S. (2017, September 18). Power of Online and Social Media: Changing the Political Landscape. *ReportGarden*. Retrieved from <https://reportgarden.com/2017/09/18/social-media-political-landscape/>

- Kim, Y. (2015). Does disagreement mitigate polarization? How selective exposure and disagreement affect political polarization. *Journalism & Mass Communication Quarterly*, 92(4). doi:10.1177/1077699015596328
- Kim, Y., Chen, H., & Gil de Zúñiga, H. (2013). Stumbling upon news on the Internet: Effects of incidental news exposure and relative entertainment use on political engagement. *Computers in Human Behavior*, 29, 2607-2614. doi:10.1016/j.chb.2013.06.005
- Kleider-Offutt, H. M., Bond, A. D., Williams, S. E., & Bohil, C. J. (2018). When a face type is perceived as threatening: Using general recognition theory to understand biased categorization of Afrocentric faces. *Memory & Cognition*, 46(5), 716-728. doi:10.3758/s13421-018-0801-0
- Knobloch-Westerwick, S., Johnson, B. K., & Westerwick, A. (2014). Confirmation bias on online searches: Impacts of selective exposure before an election on political attitude strength and shifts. *Journal of Computer-Mediated Communication*, 20(2). doi:10.1111/jcc4.12105
- Kruglanski, A. W., & Freund, T. (1983). The freezing and unfreezing of lay-inferences: Effects on impression primacy, ethnic stereotyping, and numerical anchoring. *Journal of Experimental Social Psychology*, 19, 448-468. doi:10.1016/0022-1031(83)90022-7
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480-498. doi:10.1037/0033-2909.108.3.480
- Lau, T., Pouncy, H. T., Gershman, S. J., & Cikara, M. (2018). Discovering social groups via latent structure learning. *Journal of Experimental Social Psychology: General*. Advance online publication. doi:10.1037/xge0000470
- Lee, J. K., & Kim, E. (2017). Incidental exposure to news: Predictors in the social media setting and effects on information gain online. *Computers in Human Behavior*, 75, 1008-1015. doi:10.1016/j.chb.2017.02.018
- Lemert, E. (1967). *Human deviance, social problems and social control*. Prentice-Hall.
- Levine, J. M. (1989). Reaction to opinion deviance in small groups. In P. B. Paulus (Ed.), *Psychology of group influence* (pp. 187-231). Lawrence Erlbaum Associates, Inc.
- Lim, S., & Cortina, L. M. (2005). Interpersonal mistreatment in the workplace: the interface and impact of general incivility and sexual harassment. *Journal of Applied Psychology*, 90(3), 483-496. doi:10.1037/0021-9010.90.3.483
- Liu, W., Zhao, S., Shi, L., Zhang, Z., Liu, X., Li, L., Duan, X., Li, G., Lou, F., Jia, X., Fan, L., Sun, T., & Ni, X. (2018). Workplace violence, job satisfaction, burnout, perceived organisational support and their effects on turnover intention among Chinese nurses in

- tertiary hospitals: a cross-sectional study. *BMJ Open*, 8(6). doi:10.1136/bmjopen-2017-019525
- Lord, C. G., Ross, L., Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37(11), 2098-2109. doi:10.1037/0022-3514.37.11.2098
- Lopes, B., & Yu, H. (2017). Who do you troll and Why: An investigation into the relationship between the Dark Triad Personalities and online trolling behaviours towards popular and less popular Facebook profiles. *Computers in Human Behavior*, 77, 69-76. doi:10.1016/j.chb.2017.08.036
- Luhtanen, R., & Crocker, J. (1992). A collective self-esteem scale: Self-evaluation of one's social identity. *Personality and Social Psychology Bulletin*, 18, 302-318. doi:10.1177/0146167292183006
- Maccoby, E. E. (1990). Gender as a social category. In S. Chess & M. E. Hertzig (Eds.), *Annual progress in child psychiatry and child development, 1989* (p. 127–150). Brunner/Mazel. (Reprinted from "Developmental Psychology," 1988, Vol. 24, No. 6, 755-765)
- Maltby, J., Day, L., Hatcher, R. M., Tazzyman, S., Flowe, H. D., Palmer, E. J., Frosch, C. A., O'Reilly, M., Jones, C., Buckley, C., Knieps, M., & Cutts, K. (2016). Implicit theories of online trolling: Evidence that attention-seeking conceptions are associated with increased psychological resilience. *British Journal of Psychology*, 107, 448-466. doi:10.1111/bjop.12154
- Manosevitch, E., & Walker, D. M. (2002). Reader comments to online opinion journalism: A space of public deliberation, International Symposium on Online Journalism, Austin, TX, April 17-18, 2009.
- March, E., Grieve, R., Marrington, J., & Jonason, P. K. (2017). Trolling on Tinder® (and other dating apps): Examining the role of the Dark Tetrad and impulsivity. *Personality and Individual Differences*, 110, 139-143. doi:10.1016/j.paid.2017.01.025
- Marques, J. M., Yzerbyt, V., & Leyens, J. (1988). The "Black Sheep Effect": Extremity of judgments towards ingroup members as a function of group identification. *European Journal of Social Psychology*, 18(1), 1-16. doi:10.1002/ejsp.2420180102
- Matsa, K. E., & Shearer, E. (2018, September 10). News Use Across Social Media Platforms 2018. *Pew Research Center: Journalism & Media*. Retrieved from <http://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/>
- McBride, D. M. (2020). *The process of research and statistical analysis in psychology*. SAGE Publications, Inc.

- Mikal, J. P., Rice, R. E., Kent, R. G., & Uchino, B. N. (2014). Common voice: Analysis of content convergence on a website characterized by group identification and social attraction. *Computers in Human Behavior*, *35*, 506-515. doi:10.1016/j.chb.2014.02.036
- Mikal, J. P., Rice, R. E., Kent, R. G., & Uchino, B. N. (2015). 100 million strong: A case study of group identification and deindividuation on Imgur.com. *New Media & Society*, *18*(11), 2485-2506. doi:10.1177/1461444815588766
- Mutz, D. C., & Reeves, B. (2005). The new videomalaise: Effects of televised incivility on political trust. *American Political Science Review*, *99*(1), 1-15. doi:10.1017/S0003055405051452
- Ng, E. W. J., & Detenber, B. H. (2006). The impact of synchronicity and civility in online political discussions on perceptions and intentions to participate. *Journal of Computer-Mediated Communication*, *10*(3). doi:10.1111/j.1083-6101.2005.tb00252.x
- Nicholls, S. B., & Rice, R. E. (2017). A dual-identity model of responses to deviance in online groups: Integrating social identity theory and expectancy violations theory. *Communication Theory*, *27*(3). doi:10.1111/comt.12113
- Papacharissi, Z. (2004). Democracy online: civility, politeness, and the democratic potential of online political discussion groups. *New Media & Society*, *6*(2). doi:10.1177/1461444804041444
- Penney, L. M., Spector, P. E. (2005). Job stress, incivility, and counterproductive work behavior (CWB): The moderating role of negative affectivity. *Journal of Organizational Behavior*, *26*(7), 777-796. doi:10.1002/job.336
- Phelan, J. C., Lucas, J. W., Teachman, B., Braverman, B. H., Namaky, N., & Greenberg, M. (2018). A comparison of status and stigma processes: Explicit and implicit appraisals of 'mentally ill people' and 'uneducated people'. *Stigma and Health*, *4*(2), 213-224. doi:10.1037/sah0000106
- Pyszczynski, T., & Greenberg, J. (1987). Toward and integration of cognitive and motivational perspectives on social inference: A biased hypothesis-testing model. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 297-340). Academic Press.
- Rösner, L., Winter, S., & Krämer, N. C. (2016). Dangerous minds? Effects of uncivil online comments on aggressive cognitions, emotions, and behavior. *Computers in Human Behavior*, *58*(C), 461-470. doi:10.1016/j.chb.2016.01.022
- Shachaf, P., & Hara, N. (2010). Beyond vandalism: Wikipedia trolls. *Journal of Information Science*, *36*(3), 357-370. doi:10.1177/0165551510365390

- Schilpzand, P., & Huang, L. (2018). When and how experienced incivility dissuades proactive performance: An integration of sociometer and self-identity orientation perspectives. *Journal of Applied Psychology, 103*(8), 828-841. doi:10.1037/apl0000303
- Schwartz, M. (2008, August 3). The Trolls Among Us. *The New York Times Magazine*. Retrieved from <https://www.nytimes.com/2008/08/03/magazine/03trolls-t.html>
- Scott, C. R., Rains, S. A., & Haseki, M. (2016). Anonymous communication unmasking findings across fields. *Annals of the International Communication Association, 35*(1), 299-340. doi:10.1080/23808985.2011.11679120
- Seigfried-Spellar, K. C., & Lankford, C. (2018). Personality and online environment factors differ for posters, trolls, lurkers, and confessors on Yik Yak. *Personality and Individual Differences, 124*, 54-56. doi:10.1016/j.paid.2017.11.047
- Sest, N., & March, E. (2017). Constructing the cyber-troll: Psychopathy, sadism, and empathy. *Personality and Individual Differences, 119*, 69-72. doi:10.1016/j.paid.2017.06.038
- Sherman, D. K., & Cohen, G. L. (2006). The psychology of self-defense: Self-affirmation theory. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (pp. 183-242). San Diego, CA: Academic Press. doi:10.1016/S0065-2601(06)38004-5
- Sobieraj, S., & Berry, J. M. (2011). From incivility to outrage: Political discourse in blogs, talk radio, and cable news. *Political Communication, 28*(1), 19-41. doi:10.1080/10584609.2010.542360
- Statista Research Department (2016). U.S. Reddit user distribution 2016, by gender. *Statista*. Retrieved from <https://www.statista.com/statistics/517155/reddit-user-distribution-usa-gender/>
- Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. *Advances in Experimental Social Psychology, 21*, 261-302. doi:10.1016/S0065-2601(08)60229-4
- Strauss, V. (2014, September 17). More students are illegally downloading college textbooks for free. *The Washington Post*. Retrieved from <https://www.washingtonpost.com/>
- Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science, 50*(3). doi:10.1111/j.1540-5907.2006.00214.x
- Tajfel, H. (1959). Quantitative judgment in social perception. *British Journal of Psychology, 50*, 16-29. doi:10.1111/j.2044-8295.1959.tb00677.x
- Tajfel, H., & Turner, J. C. (1979). An Integrative Theory of Intergroup Conflict. In S. Worchel (Ed.), *The Social Psychology of Intergroup Relations* (pp. 33-47). Brooks/Cole.

- Tetlock, P. E. (1983). Accountability and the perseverance of first impressions. *Social Psychology Quarterly*, 46, 285-292. doi:10.2307/3033716
- Thacker, S., & Griffiths, M. D. (2012). An exploratory study of trolling in online video gaming. *International Journal of Cyber Behavior, Psychology and Learning*, 2(4), 17-33. doi:10.4018/ijcbpl.2012100102
- Thorson, K., Vraga, E., & Ekdale, B. (2010). Credibility in context: How uncivil online commentary affects news credibility. *Mass Communication and Society*, 13(3), 289-313. doi:10.1080/15205430903225571
- Turner, J. C., Hogg, M. C., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Basil Blackwell.
- Van Eemeren, F. H., Garssen, B., & Meuffels, B. (2012). The disguised abusive ad hominem empirically investigated: Strategic manoeuvring with direct personal attacks. *Thinking & Reasoning*, 18(3), 344-364. doi:10.1080/13546783.2012.678666
- Walther, J. B., & Parks, M. R. (2002). Cues Filtered Out, Cues Filtered In: Computer-Mediated Communication and Relationships. In M. L. Knapp & J. A. Daly (Eds.) *Handbook of Interpersonal Communication* (pp. 529-563). SAGE.
- Wamsley, D. M. (2018). "Ignore them, they're a troll!": Labeling counter-attitudinal Facebook comments as trolling. Unpublished raw data.
- Wamsley, D. M. (2018). *Open-ended definitions of trolling behavior*. Unpublished raw data.
- Wang, M. Y., & Silva, D. E. (2017). A slap or a jab: An experiment on viewing uncivil political discussions on Facebook. *Computers in Human Behavior*, 81. doi:10.1016/j.chb.2017.11.041
- Weber, P., Prochazka, F., & Schweiger, W. (2017). Why user comments affect the perceived quality of journalistic content: the role of judgment processes. *Journal of Media Psychology: Theories, Methods, and Applications*. Advance online publication. doi:10.1027/1864-1105/a000217
- Webster, J. R., Adams, G. A., Maranto, C. L., Beehr, T. A. (2018). 'Dirty' workplace politics and well-being: The role of gender. *Psychology of Women Quarterly*, 42(3), 361-377. doi:10.1177/0361684318769909
- Weeks, B. E., Lane, D. S., Kim, D. H., Lee, S. S., & Kwak, N. (2017). Incidental exposure, selective exposure, and political information sharing: Integrating online exposure patterns and expression on social media. *Journal of Computer-Mediated Communication*, 22(6), 363-379. doi:10.1111/jcc4.12199

- Westen, D., Blagov, P. S., Harenski, K., Kilts, C., & Hamann, S. (2006). Neural bases of motivated reasoning: An fMRI study emotional constraints on partisan political judgment in the 2004 U.S. Presidential Election. *Journal of Cognitive Neuroscience*, *18*(11), 1947-1958. doi:10.1162/jocn.2006.18.11.1947
- Westerwick, A., Johnson, B. K., & Knobloch-Westerwick, S. (2017). Confirmation biases in selective exposure to political online information: Source bias vs. content bias. *Communication Monographs*, *84*(3), 343-364. doi:10.1080/03637751.2016.1272761
- Wojcieszak, M. E., & Mutz, D. C. (2009). Online groups and political discourse: Do online discussion spaces facilitate exposure to political disagreement? *Journal of Communication*, *59*(1). doi:10.1111/j.1460-2466.2008.01403.x
- Yang, L., & Unnava, H. R. (2016). Ambivalence, selective exposure, and negativity effect. *Psychology & Marketing*, *33*(5). doi:10.1002/mar.20878
- Yzerbyt, V. Y., Coull, A., & Rocher, S. (1999). Fencing off the deviant: The role of cognitive resources in the maintenance of stereotypes. *Journal of Personality and Social Psychology*, *76*, 377-387. doi:10.1037/0022-3514.77.3.449

Tables

Table 1

ANOVA Results – Perception of Incivility Between Subjects Effects (Manipulation Check)

Predictor	<i>df</i>	MS	<i>F</i>	<i>p</i>	η^2
Civility	1	224.39	113.70	< 0.001*	0.194
Label	2	2.21	1.12	0.33	0.004
Civ X Label	2	0.77	0.39	0.69	0.001

Note. * $p < .05$.

Table 2

ANOVA Results – Anger Between Subjects Effects

Predictor	<i>df</i>	MS	<i>F</i>	<i>p</i>	η^2
Civility	1	9.36	10.25	0.001*	0.021
Label	2	0.29	0.32	0.72	0.001
Civ X Label	2	0.35	0.38	0.69	0.002

Note. * $p < .05$.

Table 3

ANOVA Results – Anxiety Between Subjects Effects

Predictor	<i>df</i>	MS	<i>F</i>	<i>p</i>	η^2
Civility	1	0.16	0.14	0.71	0.000
Label	2	0.03	0.03	0.97	0.000
Civ X Label	2	0.90	0.81	0.45	0.003

Note. * $p < .05$.

Table 4

ANOVA Results – Attitude Certainty Between Subjects Effects

Predictor	<i>df</i>	MS	<i>F</i>	<i>p</i>	η^2
Civility	1	7.09	4.53	0.034*	0.009
Label	2	5.79	3.70	0.025*	0.015
Civ X Label	2	0.69	0.44	0.64	0.002

Note. * $p < .05$.

Table 5

Tukey's HSD Results – Attitude Certainty Comparison Between Label Conditions

Predictor	Mean Diff	SE	<i>t</i>	p_{Tukey}	Cohen's <i>d</i>
None - Rude	0.35	.14	2.47	0.037*	0.27
None - Troll	0.03	.14	0.23	0.97	0.03
Rude - Troll	-0.32	.14	-2.23	0.06 †	-0.25

Note. * $p < .05$.

Note. † Indicates marginal significance.

Table 6
ANOVA Results – Open Mindedness Between Subjects Effects

Predictor	<i>df</i>	MS	<i>F</i>	<i>p</i>	η^2
Civility	1	0.12	0.07	0.80	0.000
Label	2	3.56	1.96	0.14	0.008
Civ X Label	2	0.09	0.05	0.95	0.000

Note. **p* < .05.

Table 7
ANOVA Results – Deliberative Potential Between Subjects Effects

Predictor	<i>df</i>	MS	<i>F</i>	<i>p</i>	η^2
Civility	1	1.35	1.01	0.32	0.002
Label	2	0.62	0.46	0.63	0.002
Civ X Label	2	0.20	0.15	0.86	0.001

Note. **p* < .05.

Table 8
ANOVA Results – University Identity Between Subjects Effects

Predictor	<i>df</i>	MS	<i>F</i>	<i>p</i>	η^2
Civility	1	0.46	0.28	0.60	0.001
Label	2	2.56	1.53	0.22	0.006
Civ X Label	2	3.70	2.22	0.11	0.009

Note. **p* < .05.

Table 9
ANOVA Results – Discussion Identity Importance Between Subjects Effects

Predictor	<i>df</i>	MS	<i>F</i>	<i>p</i>	η^2
Civility	1	11.73	7.38	0.007*	0.015
Label	2	1.10	0.69	0.50	0.003
Civ X Label	2	2.69	1.69	0.18	0.007

Note. **p* < .05.

Table 10
ANOVA Results – Attitude Importance Between Subjects Effects

Predictor	<i>df</i>	MS	<i>F</i>	<i>p</i>	η^2
Civility	1	1.12	0.61	0.43	0.001
Label	2	4.61	2.53	0.08 †	0.011
Civ X Label	2	2.46	1.35	0.26	0.006

Note. **p* < .05.

Note. † Indicates marginal significance.

Table 11

Tukey's HSD Results – Attitude Importance Comparison Between Label Conditions

Predictor	Mean Diff	SE	<i>t</i>	<i>p</i> _{Tukey}	Cohen's <i>d</i>
None - Rude	-0.13	.15	-0.87	0.66	-0.10
None - Troll	-0.34	.15	-2.23	0.06 †	-0.26
Rude - Troll	-0.21	.15	-1.35	0.37	-0.16

Note. **p* < .05.

Note. † Indicates marginal significance.

Table 12

ANOVA Results – Intention to Participate Between Subjects Effects

Predictor	<i>df</i>	MS	<i>F</i>	<i>p</i>	η^2
Civility	1	13.90	5.48	0.020*	0.011
Label	2	2.15	0.85	0.43	0.004
Civ X Label	2	7.51	2.96	0.053 †	0.012

Note. **p* < .05.

Note. † Indicates marginal significance.

Table 13

Simple Effects Test Results – Difference in Intention to Participate Between Civility Condition

	<i>df</i>	MS	<i>F</i>	<i>p</i>
None	1	1.55	0.61	0.44
Rude	1	< 0.001	< 0.001	0.99
Troll	1	27.40	10.81	0.001*

Note. **p* < .05.

Table 14

Simple Effects Test Results – Difference in Intention to Participate Between Label Condition

	<i>df</i>	MS	<i>F</i>	<i>p</i>
Civil	2	1.57	0.62	0.54
Uncivil	2	8.64	3.41	0.034*

Note. **p* < .05.

Table 15

Tukey's HSD Results – Intention to Participate Comparison Between Label Conditions (Within Uncivil Condition)

Predictor	Mean Diff	SE	<i>t</i>	<i>p</i> _{Tukey}	Cohen's <i>d</i>
None - Rude	0.01	.26	0.06	0.99	0.01
None - Troll	-0.55	.25	-2.23	0.068 †	-0.36
Rude - Troll	-0.57	.25	-2.22	0.070 †	-0.36

Note. **p* < .05.

Note. † Indicates marginal significance.

Table 16
Independent Samples T-Test Results – Attitude Strength

	<i>t</i>	<i>df</i>	<i>p</i>	Cohen's <i>d</i>
Attitude Strength	-2.107	473	0.036*	-0.196

Note. * $p < .05$.

Note. Unequal cell sizes.

Table 17
ANOVA Results – Moderating Effects of TI Attitude on Anger

Predictor	<i>df</i>	MS	<i>F</i>	<i>p</i>	η^2
Civility	1	5.44	6.23	0.013*	0.013
Label	2	0.05	0.06	0.94	0.000
TI Attitude	1	15.82	18.14	< 0.001*	0.037
Civ X Label	2	0.21	0.24	0.78	0.001
Civ X TI	1	7.46	8.56	0.004*	0.017
Label X TI	2	0.99	1.14	0.32	0.005
Civ X Label X TI	2	0.11	0.12	0.89	0.000

Note. * $p < .05$.

Note. Unequal cell sizes.

Table 18
Simple Effects Test Results – Difference in Anger Between TI Attitude

	<i>df</i>	MS	<i>F</i>	<i>p</i>
Civil	1	0.97	1.12	0.29
Uncivil	1	22.39	25.67	< 0.001*

Note. * $p < .05$.

Note. Unequal cell sizes.

Table 19
Simple Effects Test Results – Difference in Anger Between Civility Condition

	<i>df</i>	MS	<i>F</i>	<i>p</i>
Anti	1	0.04	0.04	0.84
Pro	1	15.67	17.96	< 0.001*

Note. * $p < .05$.

Note. Unequal cell sizes.

Table 20

ANOVA Results – Moderating Effects of TI Attitude on Intention to Participate

Predictor	<i>df</i>	MS	<i>F</i>	<i>p</i>	η^2
Civility	1	9.43	3.83	0.051*	0.008
Label	2	1.05	0.43	0.65	0.002
TI Attitude	1	10.46	4.25	0.040*	0.009
Civ X Label	2	9.87	4.01	0.019*	0.016
Civ X TI	1	18.62	7.57	0.006*	0.015
Label X TI	2	7.58	3.08	0.047*	0.012
Civ X Label X TI	2	2.47	1.01	0.37	0.004

Note. * $p < .05$.

Note. Unequal cell sizes.

Table 21

Simple Effects Test Results – Difference in Intention to Participate Between TI Attitude

	<i>df</i>	MS	<i>F</i>	<i>p</i>
Civil	1	1.40	0.57	0.45
Uncivil	1	29.51	11.99	< 0.001*

Note. * $p < .05$.

Note. Unequal cell sizes.

Table 22

Simple Effects Test Results – Difference in Intention to Participate Between Civility Condition

	<i>df</i>	MS	<i>F</i>	<i>p</i>
Anti	1	1.02	0.42	0.52
Pro	1	35.00	14.23	< 0.001*

Note. * $p < .05$.

Note. Unequal cell sizes.

Table 23

Simple Effects Test Results – Difference in Intention to Participate Between Label Condition

	<i>df</i>	MS	<i>F</i>	<i>p</i>
Anti	2	0.68	0.28	0.76
Pro	2	9.14	3.72	0.025*

Note. * $p < .05$.

Note. Unequal cell sizes.

Table 24

Simple Effects Test Results – Difference in Intention to Participate Between TI Attitude

	<i>df</i>	MS	<i>F</i>	<i>p</i>
None	1	0.44	0.18	0.67
Rude	1	0.71	0.29	0.59
Troll	1	25.19	10.24	0.001*

Note. * $p < .05$.

Note. Unequal cell sizes.

Table 25

Tukey's HSD Results – Intention to Participate Comparison Between Label Conditions (For Pro-Trans Participants)

Predictor	Mean Diff	SE	<i>t</i>	<i>p</i> _{Tukey}	Cohen's <i>d</i>
None - Rude	0.01	.26	0.06	0.99	0.01
None - Troll	-0.55	.25	-2.23	0.068 †	-0.36
Rude - Troll	-0.57	.25	-2.22	0.070 †	-0.36

Note. * $p < .05$.

Note. † Indicates marginal significance.

Table 26

ANOVA Results – Moderating Effects of Gender on Attitude Identity Importance

Predictor	<i>df</i>	MS	<i>F</i>	<i>p</i>	η^2
Civility	1	0.00	0.00	0.97	0.000
Label	2	6.22	3.53	0.030*	0.014
Gender	1	19.53	11.08	< 0.001*	0.023
Civ X Label	2	2.43	1.38	0.25	0.006
Civ X Gender	1	5.83	3.31	0.07	0.007
Label X Gender	2	5.97	3.38	0.035*	0.014
Civ X Lab X Gen	2	0.08	0.05	0.96	0.000

Note. * $p < .05$.

Note. Unequal cell sizes.

Table 27

Simple Effects Test Results – Difference in Attitude Identity Importance Between Gender

	<i>df</i>	MS	<i>F</i>	<i>p</i>
None	1	10.43	5.92	0.015*
Rude	1	19.94	11.31	< 0.001*
Troll	1	0.02	0.01	0.92

Note. * $p < .05$.

Note. Unequal cell sizes.

Table 28

ANOVA Results – Anger Between Subjects Effects (Suspicious Removed)

Predictor	<i>df</i>	MS	<i>F</i>	<i>p</i>	η^2
Civility	1	10.179	11.09	< 0.001*	0.026
Label	2	0.90	0.98	0.38	0.005
Civ X Label	2	0.05	0.06	0.94	0.000

Note. **p* < .05.

Table 29

ANOVA Results – Anxiety Between Subjects Effects (Suspicious Removed)

Predictor	<i>df</i>	MS	<i>F</i>	<i>p</i>	η^2
Civility	1	0.06	0.05	0.83	0.000
Label	2	0.22	0.19	0.83	0.001
Civ X Label	2	0.90	0.78	0.46	0.004

Note. **p* < .05.

Table 30

ANOVA Results – Attitude Certainty Between Subjects Effects (Suspicious Removed)

Predictor	<i>df</i>	MS	<i>F</i>	<i>p</i>	η^2
Civility	1	2.27	1.45	0.23	0.003
Label	2	4.95	3.16	0.043*	0.015
Civ X Label	2	0.93	0.60	0.55	0.003

Note. **p* < .05.

Table 31

ANOVA Results – Open Mindedness Between Subjects Effects (Suspicious Removed)

Predictor	<i>df</i>	MS	<i>F</i>	<i>p</i>	η^2
Civility	1	0.36	0.20	0.66	0.000
Label	2	5.52	3.03	0.049*	0.014
Civ X Label	2	0.53	0.29	0.75	0.001

Note. **p* < .05.

Table 32

Tukey's HSD Results – Open Mindedness Comparison Between Label Conditions

Predictor	Mean Diff	SE	<i>t</i>	<i>p</i> _{Tukey}	Cohen's <i>d</i>
None - Rude	-0.16	.16	-0.99	0.58	-0.12
None - Troll	-0.39	.16	-2.45	0.039*	-0.29
Rude - Troll	-0.23	.16	-1.43	0.33	-0.17

Note. **p* < .05.

Table 33

ANOVA Results – Deliberative Potential Between Subjects Effects (Suspicious Removed)

Predictor	<i>df</i>	MS	<i>F</i>	<i>p</i>	η^2
Civility	1	0.95	0.71	0.40	0.002
Label	2	0.74	0.55	0.58	0.003
Civ X Label	2	0.12	0.09	0.92	0.000

Note. * $p < .05$.

Table 34

ANOVA Results – Intention to Participate Between Subjects Effects (Suspicious Removed)

Predictor	<i>df</i>	MS	<i>F</i>	<i>p</i>	η^2
Civility	1	8.99	3.55	0.060 †	0.008
Label	2	1.25	0.49	0.61	0.002
Civ X Label	2	5.63	2.22	0.11	0.010

Note. * $p < .05$.

Note. † Indicates marginal significance.

Table 35

ANOVA Results – University Identity Between Subjects Effects (Suspicious Removed)

Predictor	<i>df</i>	MS	<i>F</i>	<i>p</i>	η^2
Civility	1	0.80	0.50	0.48	0.001
Label	2	2.03	1.27	0.28	0.006
Civ X Label	2	3.25	2.04	0.13	0.010

Note. * $p < .05$.

Table 36

ANOVA Results – Discussion Identity Between Subjects Effects (Suspicious Removed)

Predictor	<i>df</i>	MS	<i>F</i>	<i>p</i>	η^2
Civility	1	15.36	9.65	0.002*	0.022
Label	2	0.29	0.18	0.84	0.001
Civ X Label	2	2.06	1.30	0.28	0.006

Note. * $p < .05$.

Table 37

ANOVA Results – Attitude Identity Between Subjects Effects (Suspicious Removed)

Predictor	<i>df</i>	MS	<i>F</i>	<i>p</i>	η^2
Civility	1	3.78	2.12	0.15	0.005
Label	2	2.84	1.59	0.20	0.008
Civ X Label	2	1.89	1.06	0.35	0.005

Note. * $p < .05$.

Table 38

ANOVA Results – Perception of Incivility Between Subjects Effects (Suspicious Removed)

Predictor	<i>df</i>	MS	<i>F</i>	<i>p</i>	η^2
Civility	1	219.15	112.31	< 0.001*	0.212
Label	2	0.29	0.15	0.86	0.001
Civ X Label	2	0.24	0.12	0.88	0.000

Note. * $p < .05$.

Table 39

Correlation Matrix – All Dependent Variables

	Att Certain	Open- Mind	Delib Potent	Univ Ident	Disc Ident	Att Ident	Anger	Anxiety	Intent to Part
Att Certain	---								
Open- Mind	0.398*	---							
Delib Potent	0.328*	0.515*	---						
Univ Ident	0.124*	0.151*	0.134*	---					
Disc Ident	0.339*	0.298*	0.274*	0.176*	---				
Att Ident	0.215*	0.016	0.125*	0.086	0.399*	---			
Anger	0.234*	-0.007	-0.088	0.004	0.196*	0.174*	---		
Anxiety	0.133*	-0.028	-0.082	0.011	0.081	0.166*	0.330*	---	
Intent to Part	0.387*	0.258*	0.339*	0.129*	0.309*	0.213*	0.163*	0.127*	---

Note. * $p < .05$.

Table 40

Principal Component Analysis – Factor Loadings of Each Dependent Variable

	Factor #1	Factor #2
Anger	--	0.789
Anxiety	--	0.743
Attitude Certainty	0.588	--
Open-Mindedness	0.825	--
Deliberative Potential	0.854	--
Intention to Participate	0.535	--
Discussion Identity	0.505	--
Attitude Identity	--	0.534
University Identity	--	--

Table 41
A Priori Hypotheses and Results from Primary Analysis

Hypothesis #	Hypothesis	Result
H1a	Uncivil condition → higher ratings of anger and anxiety	<i>Partially supported</i> *Supported for anger, but not anxiety
H1b	Within uncivil condition → lower ratings of anger and anxiety in the troll condition	<i>Not supported</i>
H2a	Uncivil condition → higher ratings of attitude certainty	<i>Supported</i>
H2b	Within uncivil condition → lower ratings of attitude certainty in the troll condition	<i>Not supported</i>
H3a	Uncivil condition → lower ratings of open mindedness	<i>Not supported</i>
H3b	Within uncivil condition → higher ratings of open mindedness in the troll condition	<i>Not supported</i>
H4a	Uncivil condition → higher ratings of intention to participate	<i>Supported</i>
H4b	Within uncivil condition → higher ratings of intention to participate in the troll condition	<i>Partially supported</i> *Effect was marginally significant
H5a	Uncivil condition → lower ratings of perceptions of discussion's deliberative potential	<i>Not supported</i>
H5b	Within uncivil condition → lower ratings of perceptions of discussion's deliberative potential in the troll condition	<i>Not supported</i>
H6a	No label → lower ratings of group identification than rude or troll condition	<i>Partially supported</i> *Effect was marginally significant for attitude towards transgender identities group membership
H6b	Rude label → lower ratings of group identification than troll condition	<i>Not supported</i>

Table 42

Order of discussion comments presented to participants – CIVIL & SUPPORTIVE OF TRANS IDENTITIES.

Group Member	No Label	“Rude” Label	“Troll” Label
Pro-attitudinal #1 – Dark Blue	I guess I’ll get us started... I don’t know many trans people, but I support their right to be whatever they want because they aren’t hurting anybody. That’s what our country is founded on.	I guess I’ll get us started... I don’t know many trans people, but I support their right to be whatever they want because they aren’t hurting anybody. That’s what our country is founded on.	I guess I’ll get us started... I don’t know many trans people, but I support their right to be whatever they want because they aren’t hurting anybody. That’s what our country is founded on.
Pro-attitudinal #2 - Green	I wholeheartedly agree. When I was growing up, I was taught that you are either a boy or a girl, and there was nothing you could do about that. I’ve grown, and I’m learning that it doesn’t have to be our reality. Trans people just want to be happy, and who are we to deny them that?	I wholeheartedly agree. When I was growing up, I was taught that you are either a boy or a girl, and there was nothing you could do about that. I’ve grown, and I’m learning that it doesn’t have to be our reality. Trans people just want to be happy, and who are we to deny them that?	I wholeheartedly agree. When I was growing up, I was taught that you are either a boy or a girl, and there was nothing you could do about that. I’ve grown, and I’m learning that it doesn’t have to be our reality. Trans people just want to be happy, and who are we to deny them that?
Pro-attitudinal #3 - Red	I don’t really get the whole trans bathroom scare. There are stalls in bathrooms for a reason. Noone cares what body parts you have in the bathroom. Just wash your hands when you’re done.	I don’t really get the whole trans bathroom scare. There are stalls in bathrooms for a reason. Noone cares what body parts you have in the bathroom. Just wash your hands when you’re done.	I don’t really get the whole trans bathroom scare. There are stalls in bathrooms for a reason. Noone cares what body parts you have in the bathroom. Just wash your hands when you’re done.
Pro-attitudinal #4 - Orange	Right??? just let people be who they want to be. everyone deserves equal civil rights in the United States.	Right??? just let people be who they want to be. everyone deserves equal civil rights in the United States.	Right??? just let people be who they want to be. everyone deserves equal civil rights in the United States.
Participant - Teal	N/A	N/A	N/A
Counter-attitudinal - Yellow	If I recall correctly, I learned in biology class that there are only two genders, and you can’t change your DNA. I’m not against transgender people, but I humbly believe that it’s just not natural.	If I recall correctly, I learned in biology class that there are only two genders, and you can’t change your DNA. I’m not against transgender people, but I humbly believe that it’s just not natural.	If I recall correctly, I learned in biology class that there are only two genders, and you can’t change your DNA. I’m not against transgender people, but I humbly believe that it’s just not natural.
Pro-attitudinal #5 - Black	I also remember taking biology in school. Science suggests that gender is a spectrum separate from sex/dna. Why do you even care how others identify?	I also remember taking biology in school. Science suggests that gender is a spectrum separate from sex/dna. Why do you even care how others identify?	I also remember taking biology in school. Science suggests that gender is a spectrum separate from sex/dna. Why do you even care how others identify?
Pro-attitudinal #6 - Pink	Seriously, stop the hate and misinformation. Anyone who is afraid of trans people has got it all wrong. There is nothing to be afraid of.	Seriously, stop the hate and misinformation. Anyone who is afraid of trans people has got it all wrong. There is nothing to be afraid of.	Seriously, stop the hate and misinformation. Anyone who is afraid of trans people has got it all wrong. There is nothing to be afraid of.
Pro-attitudinal #7 - Gray	It’s unfortunate how many people don’t have basic compassion for fellow humanity. My heart aches for transgender people suffering from bigotry and hate right now.	It’s unfortunate how many people don’t have basic compassion for fellow humanity. My heart aches for transgender people suffering from bigotry and hate right now.	It’s unfortunate how many people don’t have basic compassion for fellow humanity. My heart aches for transgender people suffering from bigotry and hate right now.
Pro-attitudinal #8 – Purple	Guys, the yellow commenter clearly doesn’t agree with us.	Guys, the yellow commenter is clearly being rude.	Guys, the yellow commenter is clearly being a troll.

Table 43

Order of discussion comments presented to participants – UNCIVIL & SUPPORTIVE OF TRANS IDENTITIES.

Group Member	No Label	“Rude” Label	“Troll” Label
Pro-attitudinal #1 – Dark Blue	I guess I’ll get us started... I don’t know many trans people, but I support their right to be whatever they want because they aren’t hurting anybody. That’s what our country is founded on.	I guess I’ll get us started... I don’t know many trans people, but I support their right to be whatever they want because they aren’t hurting anybody. That’s what our country is founded on.	I guess I’ll get us started... I don’t know many trans people, but I support their right to be whatever they want because they aren’t hurting anybody. That’s what our country is founded on.
Pro-attitudinal #2 - Green	I wholeheartedly agree. When I was growing up, I was taught that you are either a boy or a girl, and there was nothing you could do about that. I’ve grown, and I’m learning that it doesn’t have to be our reality. Trans people just want to be happy, and who are we to deny them that?	I wholeheartedly agree. When I was growing up, I was taught that you are either a boy or a girl, and there was nothing you could do about that. I’ve grown, and I’m learning that it doesn’t have to be our reality. Trans people just want to be happy, and who are we to deny them that?	I wholeheartedly agree. When I was growing up, I was taught that you are either a boy or a girl, and there was nothing you could do about that. I’ve grown, and I’m learning that it doesn’t have to be our reality. Trans people just want to be happy, and who are we to deny them that?
Pro-attitudinal #3 - Red	I don’t really get the whole trans bathroom scare. There are stalls in bathrooms for a reason. Noone cares what body parts you have in the bathroom. Just wash your hands when you’re done.	I don’t really get the whole trans bathroom scare. There are stalls in bathrooms for a reason. Noone cares what body parts you have in the bathroom. Just wash your hands when you’re done.	I don’t really get the whole trans bathroom scare. There are stalls in bathrooms for a reason. Noone cares what body parts you have in the bathroom. Just wash your hands when you’re done.
Pro-attitudinal #4 - Orange	Right??? just let people be who they want to be. everyone deserves equal civil rights in the United States.	Right??? just let people be who they want to be. everyone deserves equal civil rights in the United States.	Right??? just let people be who they want to be. everyone deserves equal civil rights in the United States.
Participant - Teal	N/A	N/A	N/A
Counter-attitudinal - Yellow	You uneducated liberals must have flunked basic biology. Science says there are only TWO GENDERS. Once a man, always a man!! It's been proven, assholes!!!	You uneducated liberals must have flunked basic biology. Science says there are only TWO GENDERS. Once a man, always a man!! It's been proven, assholes!!!	You uneducated liberals must have flunked basic biology. Science says there are only TWO GENDERS. Once a man, always a man!! It's been proven, assholes!!!
Pro-attitudinal #5 - Black	I also remember taking biology in school. Science suggests that gender is a spectrum. Why do you even care how others identify?	I also remember taking biology in school. Science suggests that gender is a spectrum. Why do you even care how others identify?	I also remember taking biology in school. Science suggests that gender is a spectrum. Why do you even care how others identify?
Pro-attitudinal #6 - Pink	Seriously, stop the hate and misinformation. Anyone who is afraid of trans people has got it all wrong. There is nothing to be afraid of.	Seriously, stop the hate and misinformation. Anyone who is afraid of trans people has got it all wrong. There is nothing to be afraid of.	Seriously, stop the hate and misinformation. Anyone who is afraid of trans people has got it all wrong. There is nothing to be afraid of.
Pro-attitudinal #7 - Gray	It’s unfortunate how many people don’t have basic compassion for fellow humanity. My heart aches for transgender people suffering from bigotry and hate right now.	It’s unfortunate how many people don’t have basic compassion for fellow humanity. My heart aches for transgender people suffering from bigotry and hate right now.	It’s unfortunate how many people don’t have basic compassion for fellow humanity. My heart aches for transgender people suffering from bigotry and hate right now.
Pro-attitudinal #8 – Purple	Guys, the yellow commenter clearly doesn’t agree with us.	Guys, the yellow commenter is clearly being rude.	Guys, the yellow commenter is clearly being a troll.

Table 44

Order of discussion comments presented to participants – CIVIL & NOT SUPPORTIVE OF TRANS IDENTITIES.

Group Member	No Label	“Rude” Label	“Troll” Label
Pro-attitudinal #1 – Dark Blue	I guess I’ll get us started... I’m not against transgendered people, but I believe God made only two genders. I just don’t think switching your gender later in life makes any sense.	I guess I’ll get us started... I’m not against transgendered people, but I believe God made two genders. I just don’t think switching your gender later in life makes any sense.	I guess I’ll get us started... I’m not against transgendered people, but I believe God made two genders. I just don’t think switching your gender later in life makes any sense.
Pro-attitudinal #2 - Green	I wholeheartedly agree. People should have the freedom to be whatever they want, but don’t force me to accept it. You do you, and I’ll do me.	I wholeheartedly agree. People should have the freedom to be whatever they want, but don’t force me to accept it. You do you, and I’ll do me.	I wholeheartedly agree. People should have the freedom to be whatever they want, but don’t force me to accept it. You do you, and I’ll do me.
Pro-attitudinal #3 - Red	I think that the transgender lifestyle may be confusing to young children. Sometimes, I don’t think people consider how kids might be affected by these issues.	I think that the transgender lifestyle may be confusing to young children. Sometimes, I don’t think people consider how kids might be affected by these issues.	I think that the transgender lifestyle may be confusing to young children. Sometimes, I don’t think people consider how kids might be affected by these issues.
Pro-attitudinal #4 - Orange	right??? I recently learned that being transgender is technically considered a mental illness that can have really negative effects in the long run. i truly care about the young people who are embracing this lifestyle without understanding what it could cause.	right??? I recently learned that being transgender is technically considered a mental illness that can have really negative effects in the long run. i truly care about the young people who are embracing this lifestyle without understanding what it could cause.	right??? I recently learned that being transgender is technically considered a mental illness that can have really negative effects in the long run. i truly care about the young people who are embracing this lifestyle without understanding what it could cause.
Participant - Teal	N/A	N/A	N/A
Counter-attitudinal - Yellow	It’s not a mental illness. Mental illnesses harm people... being trans isn’t harmful, it’s all the hate that people send their way that harms them. Just mind your own business.	It’s not a mental illness. Mental illnesses harm people... being trans isn’t harmful, it’s all the hate that people send their way that harms them. Just mind your own business.	It’s not a mental illness. Mental illnesses harm people... being trans isn’t harmful, it’s all the hate that people send their way that harms them. Just mind your own business.
Pro-attitudinal #5 - Black	How are we supposed to mind our business when the push for “equal rights” could potentially put others in danger? I’m sincerely concerned.	How are we supposed to mind our business when the push for “equal rights” could potentially put others in danger? I’m sincerely concerned.	How are we supposed to mind our business when the push for “equal rights” could potentially put others in danger? I’m sincerely concerned.
Pro-attitudinal #6 - Pink	On top of that, didn’t you guys hear about the man who dressed in women’s clothing to go into a dressing room? It’s starting to happen... I just don’t want that to happen to anyone on our campus.	On top of that, didn’t you guys hear about the man who dressed in women’s clothing to go into a dressing room? It’s starting to happen... I just don’t want that to happen to anyone on our campus.	On top of that, didn’t you guys hear about the man who dressed in women’s clothing to go into a dressing room? It’s starting to happen... I just don’t want that to happen to anyone on our campus.
Pro-attitudinal #7 - Gray	Mental illness or not, we need to protect women when they are in the women’s restrooms and dressing rooms Their safety comes first.	Mental illness or not, we need to protect women when they are in the women’s restrooms and dressing rooms Their safety comes first.	Mental illness or not, we need to protect women when they are in the women’s restrooms and dressing rooms Their safety comes first.
Pro-attitudinal #8 – Purple	Guys, the yellow commenter clearly doesn’t agree with us.	Guys, the yellow commenter is clearly being rude.	Guys, the yellow commenter is clearly being a troll.

Table 45

Order of discussion comments presented to participants – CIVIL & NOT SUPPORTIVE OF TRANS IDENTITIES.

Group Member	No Label	“Rude” Label	“Troll” Label
Pro-attitudinal #1 – Dark Blue	I guess I’ll get us started... I’m not against transgendered people, but I believe God made only two genders. I just don’t think switching your gender later in life makes any sense.	I guess I’ll get us started... I’m not against transgendered people, but I believe God made two genders. I just don’t think switching your gender later in life makes any sense.	I guess I’ll get us started... I’m not against transgendered people, but I believe God made two genders. I just don’t think switching your gender later in life makes any sense.
Pro-attitudinal #2 - Green	I wholeheartedly agree. People should have to freedom to be whatever they want, but don’t force me to accept it. You do you, and I’ll do me.	I wholeheartedly agree. People should have to freedom to be whatever they want, but don’t force me to accept it. You do you, and I’ll do me.	I wholeheartedly agree. People should have to freedom to be whatever they want, but don’t force me to accept it. You do you, and I’ll do me.
Pro-attitudinal #3 - Red	I think that the transgender lifestyle may be confusing to young children. Sometimes, I don’t think people consider how kids might be affected by these issues.	I think that the transgender lifestyle may be confusing to young children. Sometimes, I don’t think people consider how kids might be affected by these issues.	I think that the transgender lifestyle may be confusing to young children. Sometimes, I don’t think people consider how kids might be affected by these issues.
Pro-attitudinal #4 - Orange	right??? I recently learned that being transgender is technically considered a mental illness that can have really negative effects in the long run. i truly care about the young people who are embracing this lifestyle without understanding what it could cause.	right??? I recently learned that being transgender is technically considered a mental illness that can have really negative effects in the long run. i truly care about the young people who are embracing this lifestyle without understanding what it could cause.	right??? I recently learned that being transgender is technically considered a mental illness that can have really negative effects in the long run. i truly care about the young people who are embracing this lifestyle without understanding what it could cause.
Participant - Teal	N/A	N/A	N/A
Counter-attitudinal - Yellow	Mental illness? YEAH RIGHT. Mental illnesses harm people... being trans isn’t harmful. It’s the intolerant conservative ASSHOLES spewing all the hate that harms them. Mind your own damn business!!!	Mental illness? YEAH RIGHT. Mental illnesses harm people... being trans isn’t harmful. It’s the intolerant conservative ASSHOLES spewing all the hate that harms them. Mind your own damn business!!!	Mental illness? YEAH RIGHT. Mental illnesses harm people... being trans isn’t harmful. It’s the intolerant conservative ASSHOLES spewing all the hate that harms them. Mind your own damn business!!!
Pro-attitudinal #5 - Black	How are we supposed to mind our business when the push for “equal rights” could potentially put others in danger? I’m sincerely concerned.	How are we supposed to mind our business when the push for “equal rights” could potentially put others in danger? I’m sincerely concerned.	How are we supposed to mind our business when the push for “equal rights” could potentially put others in danger? I’m sincerely concerned.
Pro-attitudinal #6 - Pink	On top of that, didn’t you guys hear about the man who dressed in women’s clothing to go into a dressing room? It’s starting to happen I just don’t want that to happen to anyone on our campus.	On top of that, didn’t you guys hear about the man who dressed in women’s clothing to go into a dressing room? It’s starting to happen I just don’t want that to happen to anyone on our campus.	On top of that, didn’t you guys hear about the man who dressed in women’s clothing to go into a dressing room? It’s starting to happen I just don’t want that to happen to anyone on our campus.
Pro-attitudinal #7 - Gray	Mental illness or not, we need to protect women when they are in the women’s restrooms and dressing rooms. Their safety comes first.	Mental illness or not, we need to protect women when they are in the women’s restrooms and dressing rooms. Their safety comes first.	Mental illness or not, we need to protect women when they are in the women’s restrooms and dressing rooms. Their safety comes first.
Pro-attitudinal #8 – Purple	Guys, the yellow commenter clearly doesn’t agree with us.	Guys, the yellow commenter is clearly being rude.	Guys, the yellow commenter is clearly being a troll.

Figures

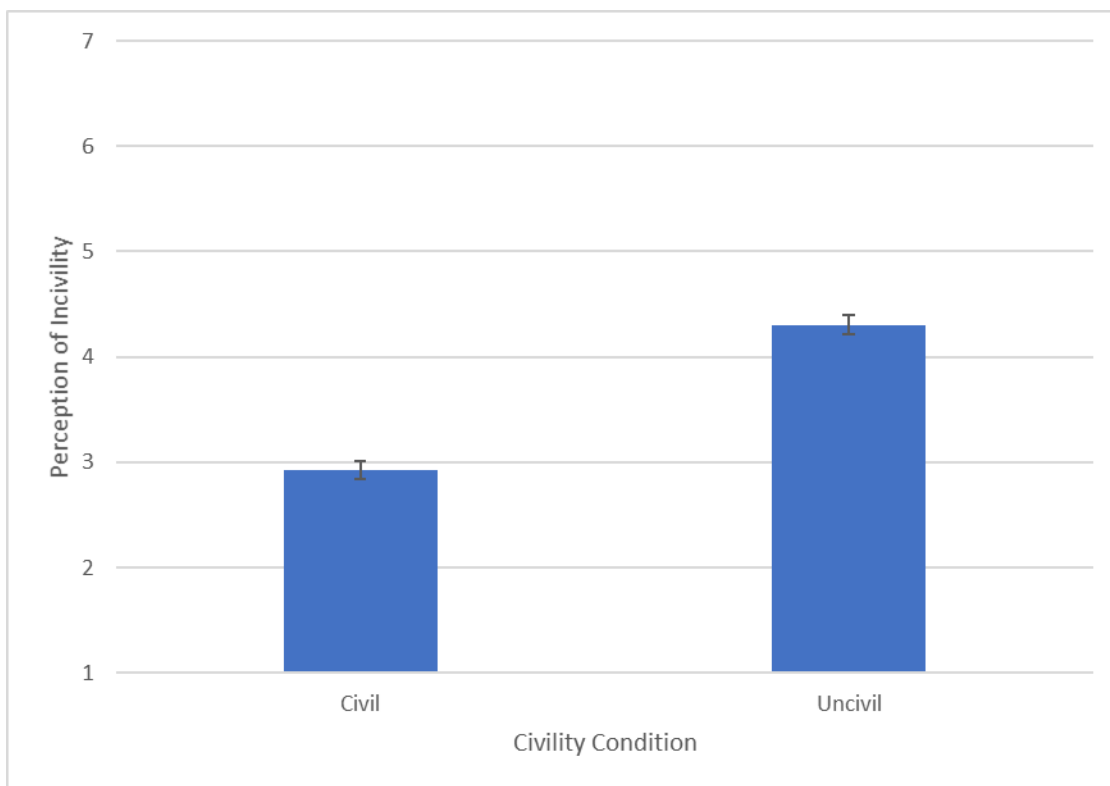


Figure 1. Perception of discussion incivility by civility condition, error bars +/-1 SE.

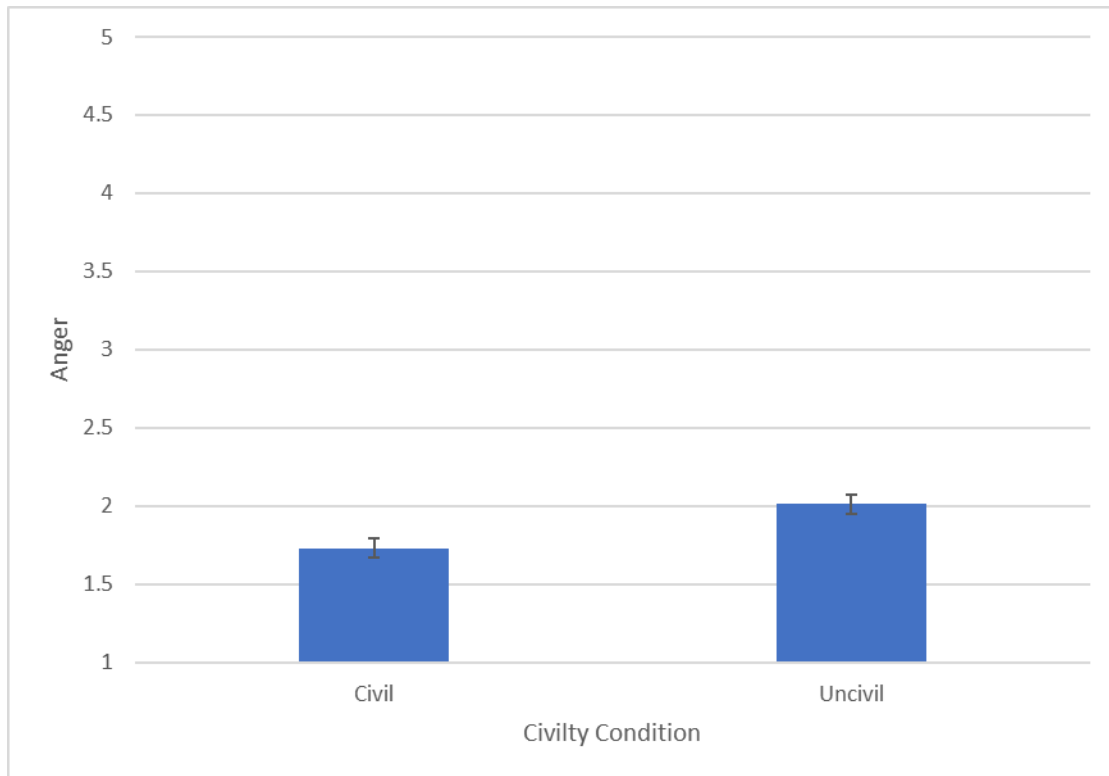


Figure 2. Ratings of anger by civility condition, error bars ± 1 SE.

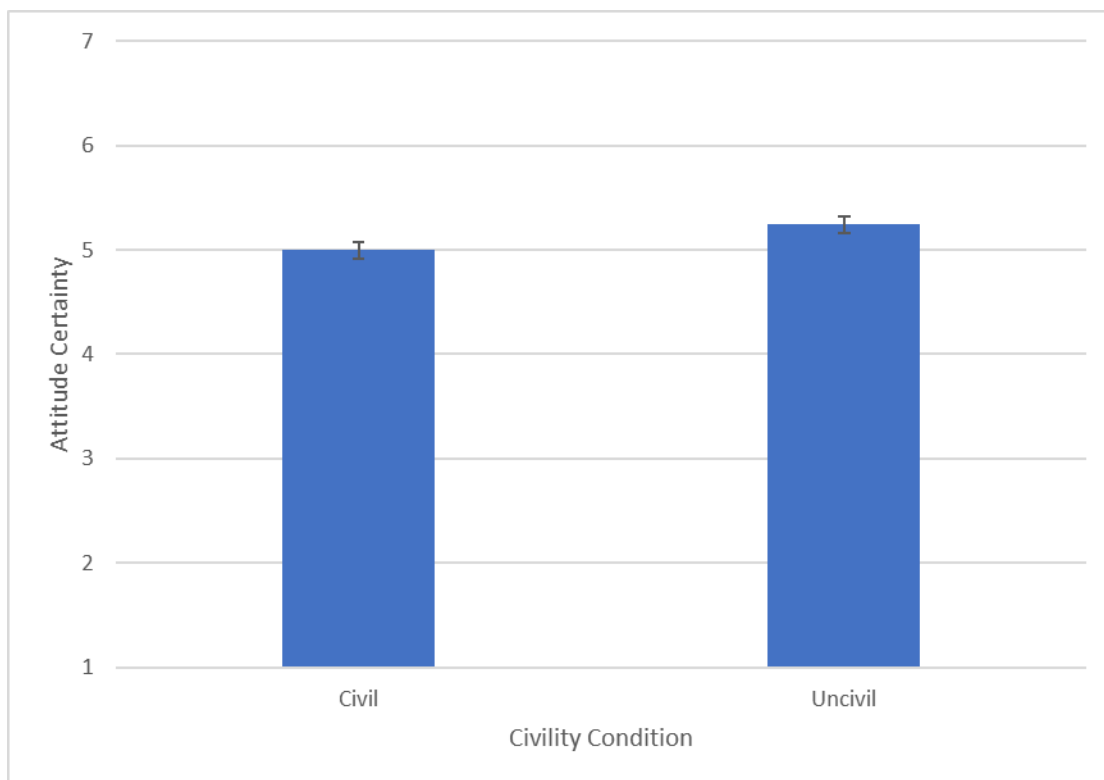


Figure 3. Ratings of attitude certainty by civility condition, error bars +/-1 SE.

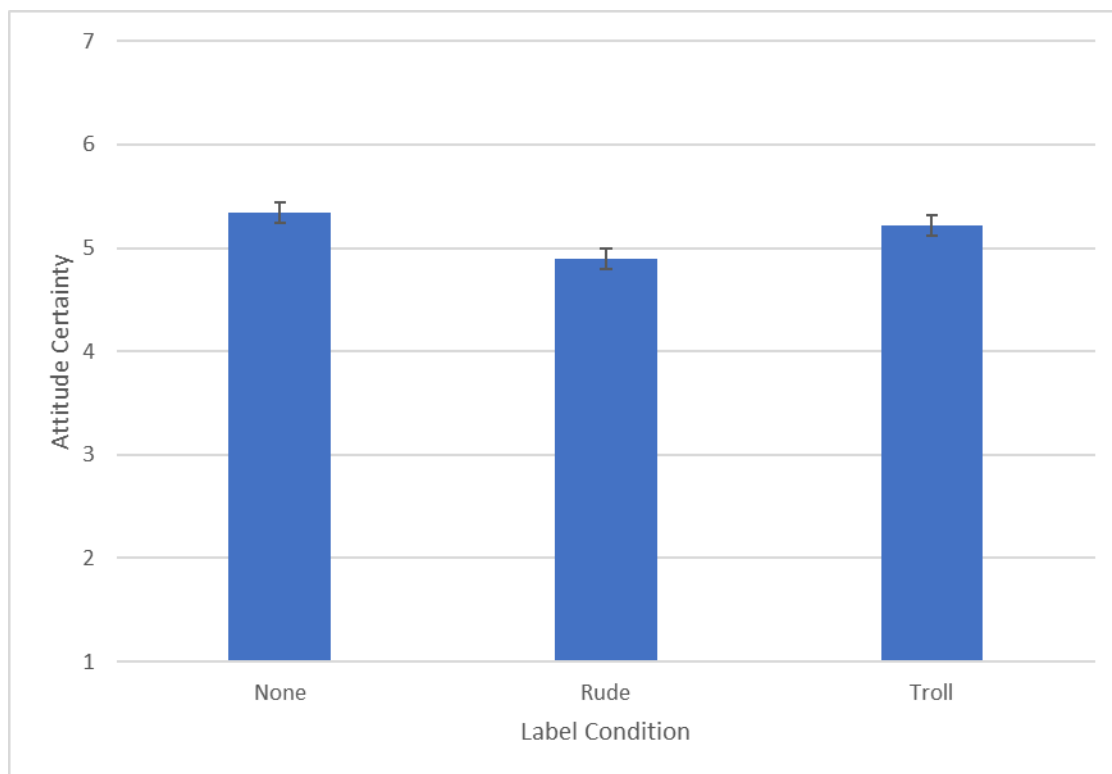


Figure 4. Ratings of attitude certainty by label condition, error bars +/-1 SE.

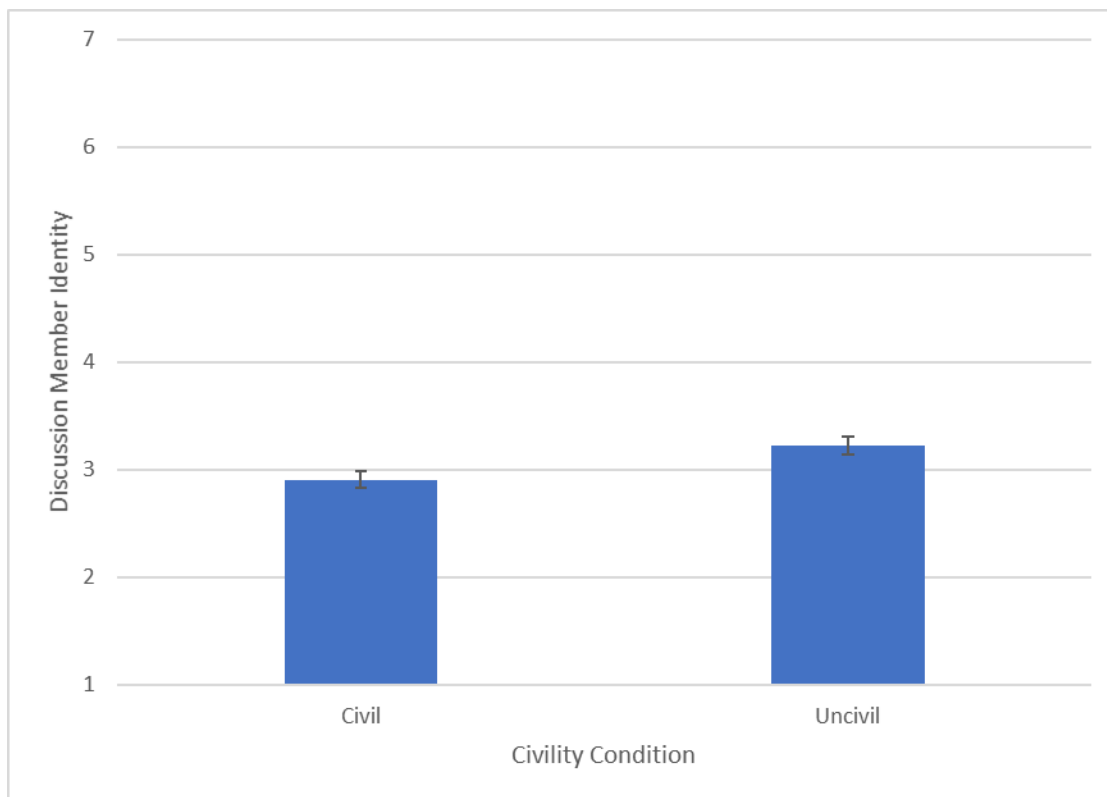


Figure 5. Ratings of discussion member identification by civility condition, error bars +/-1 SE.

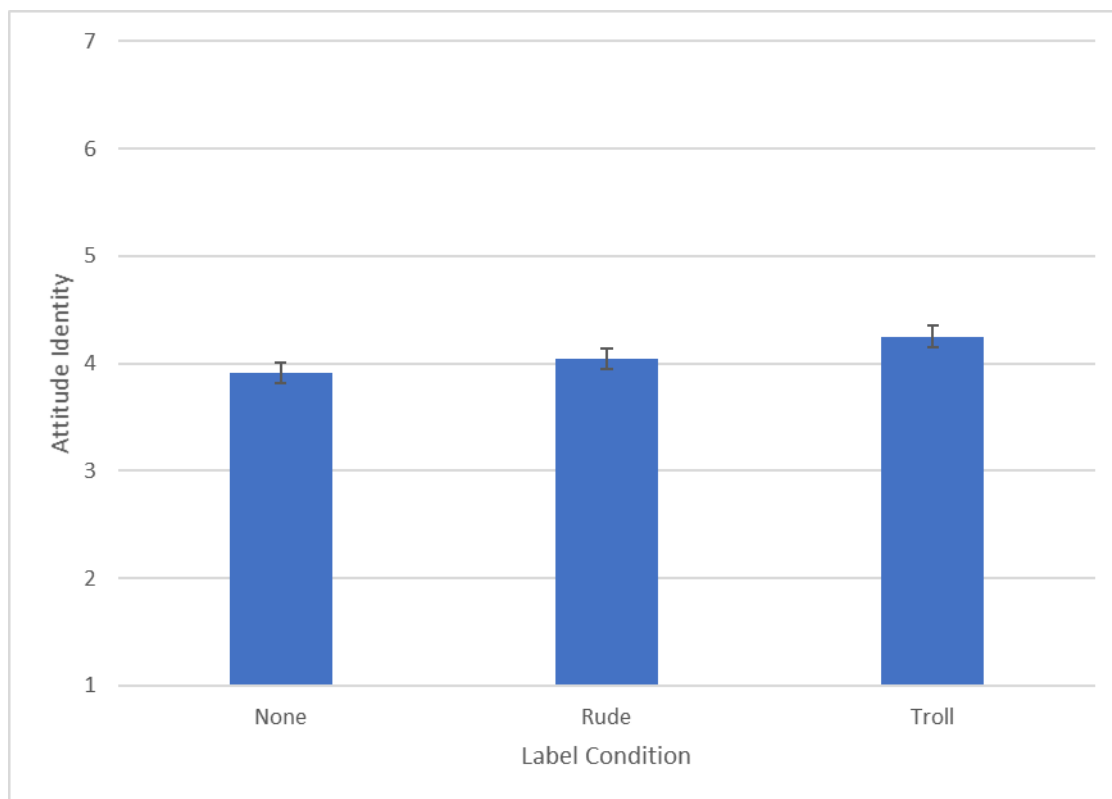


Figure 6. Ratings of pre-existing attitude identification by label condition, error bars +/-1 SE.

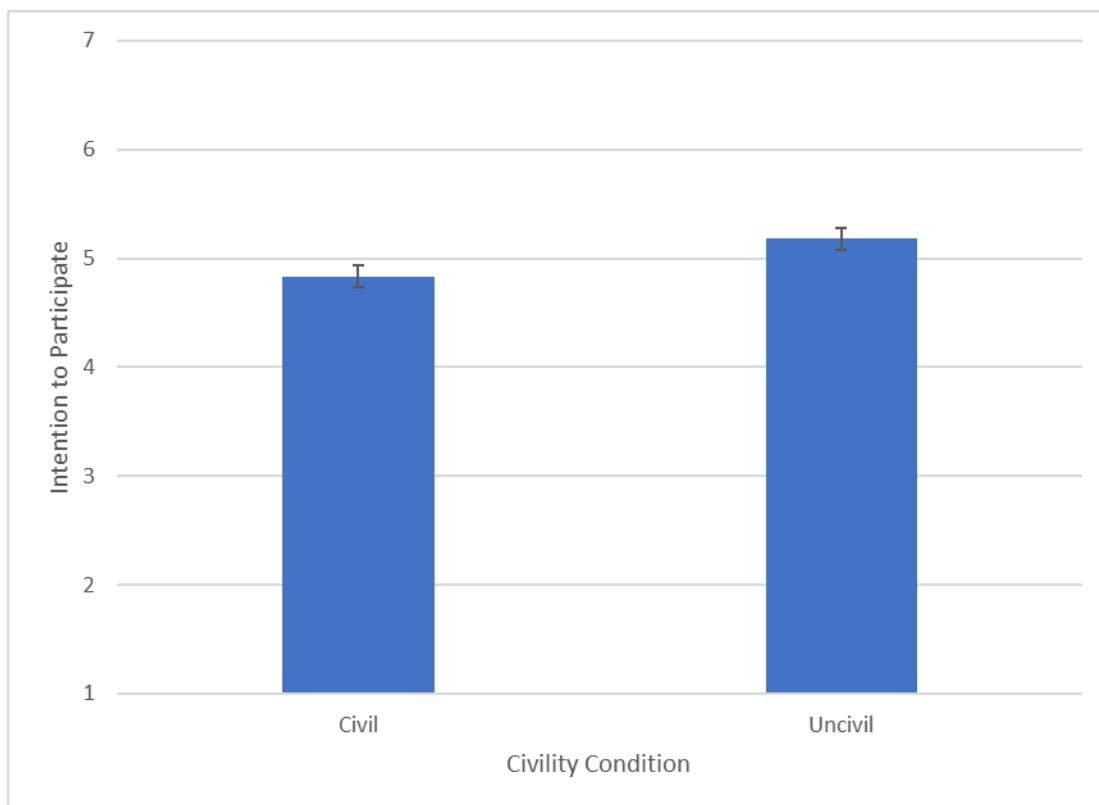


Figure 7. Ratings of participation intention by civility condition, error bars +/-1 SE.

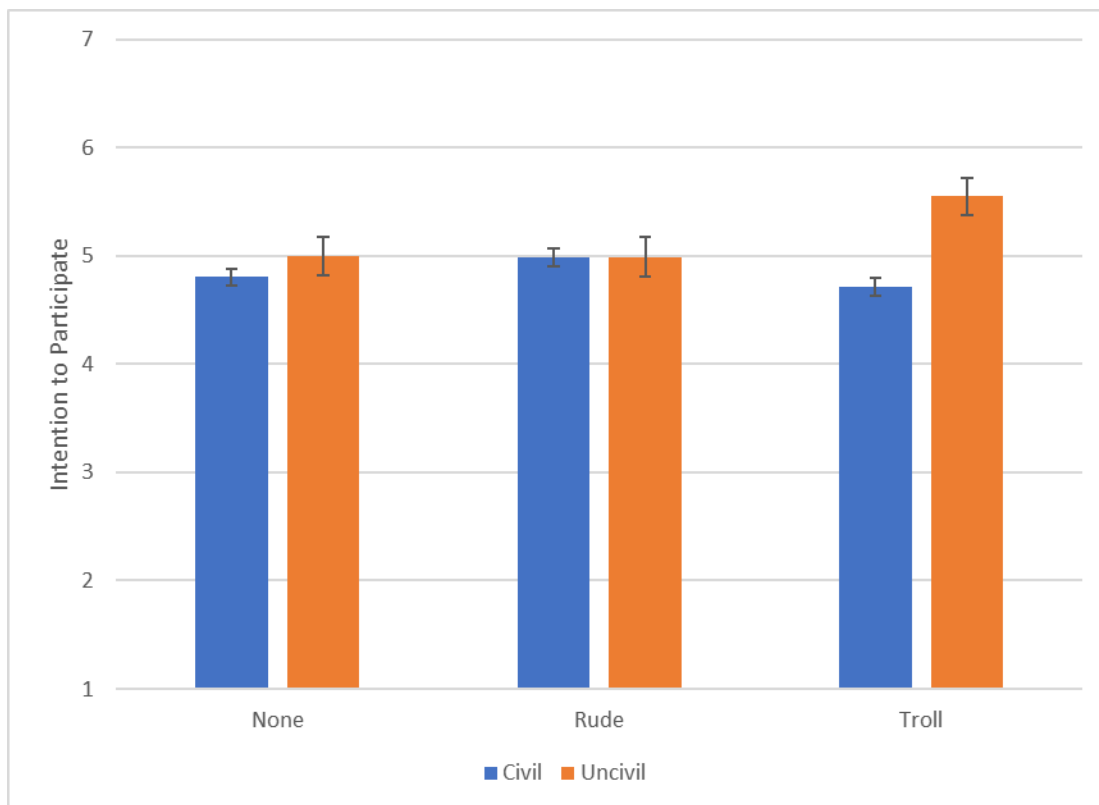


Figure 8. Ratings of participation intention by civility X label conditions, error bars +/-1 SE.

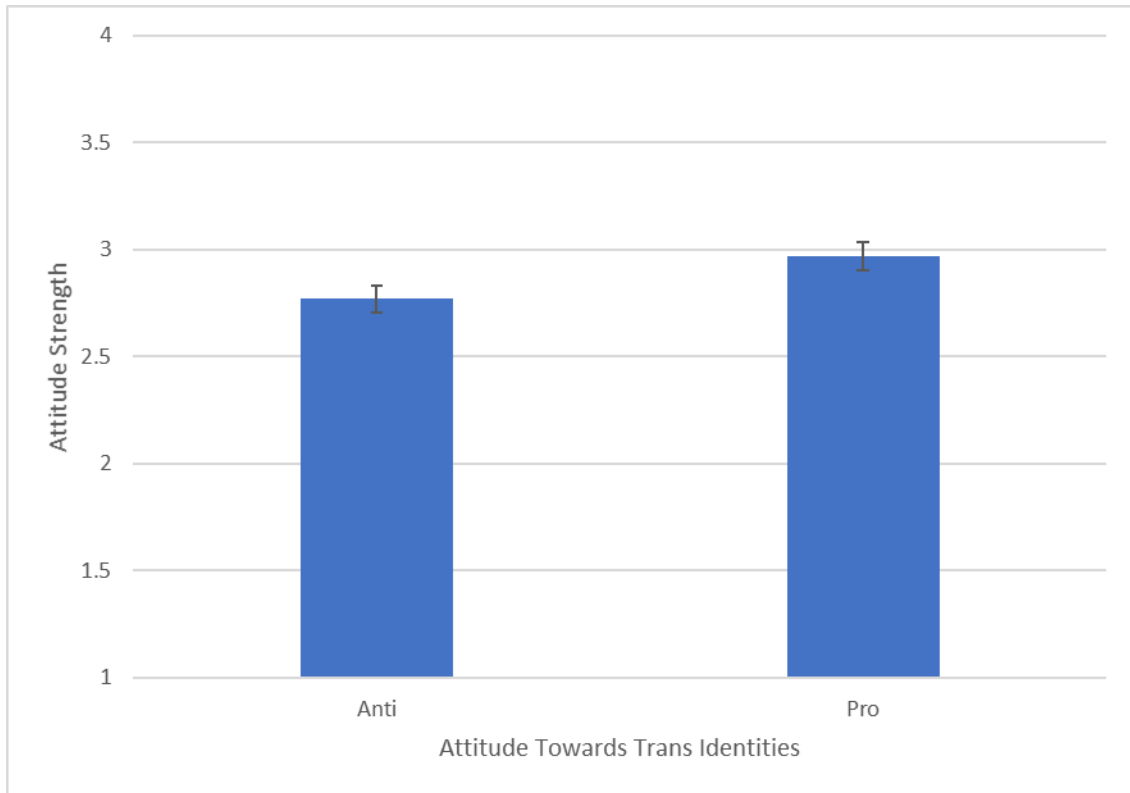


Figure 9. Ratings of attitude strength by transgender identities attitude, error bars +/-1 SE.

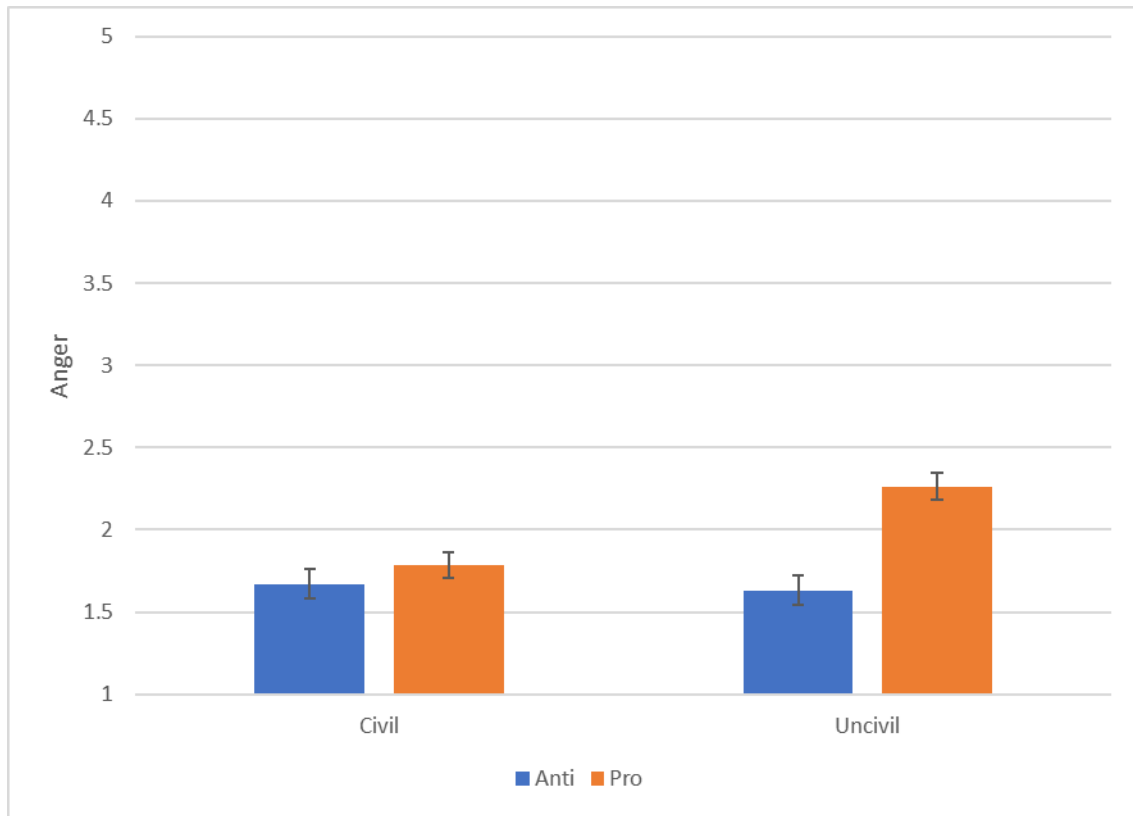


Figure 10. Ratings of anger by transgender identities attitude X civility condition, error bars ± 1 SE.

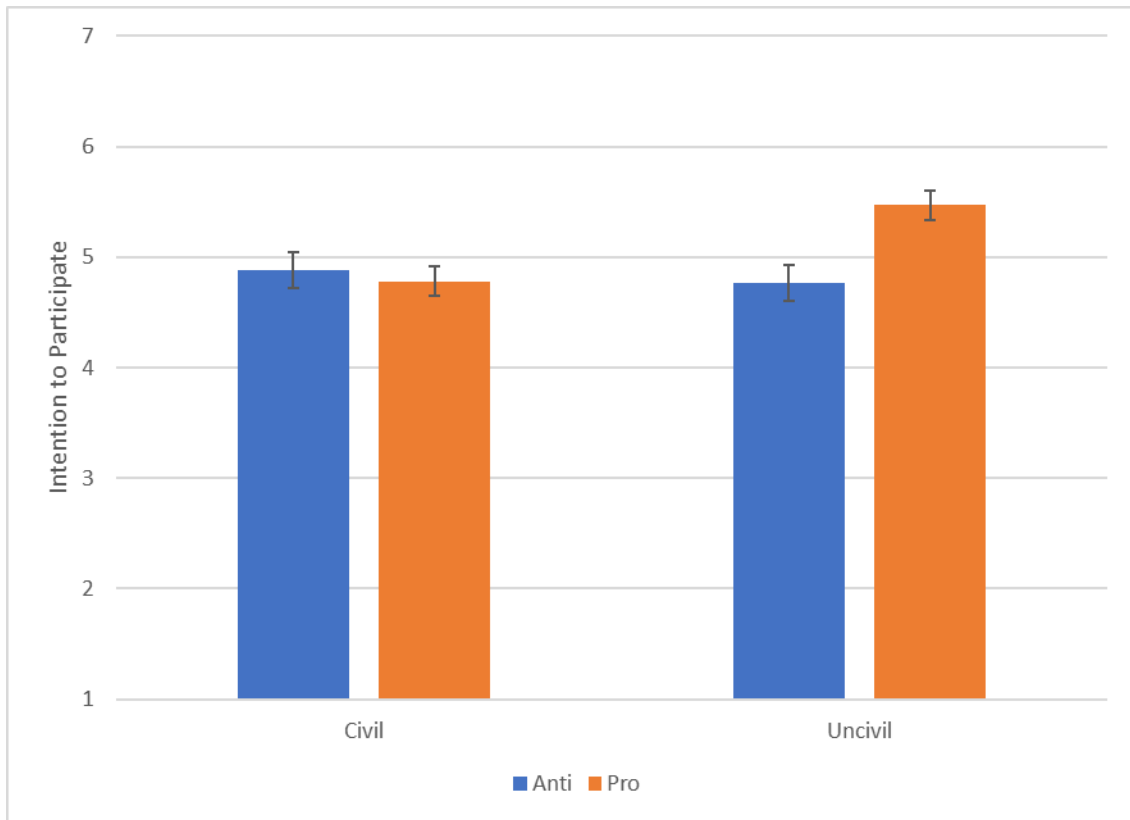


Figure 11. Ratings of participation intention by transgender identity attitude X civility condition, error bars +/-1 SE.

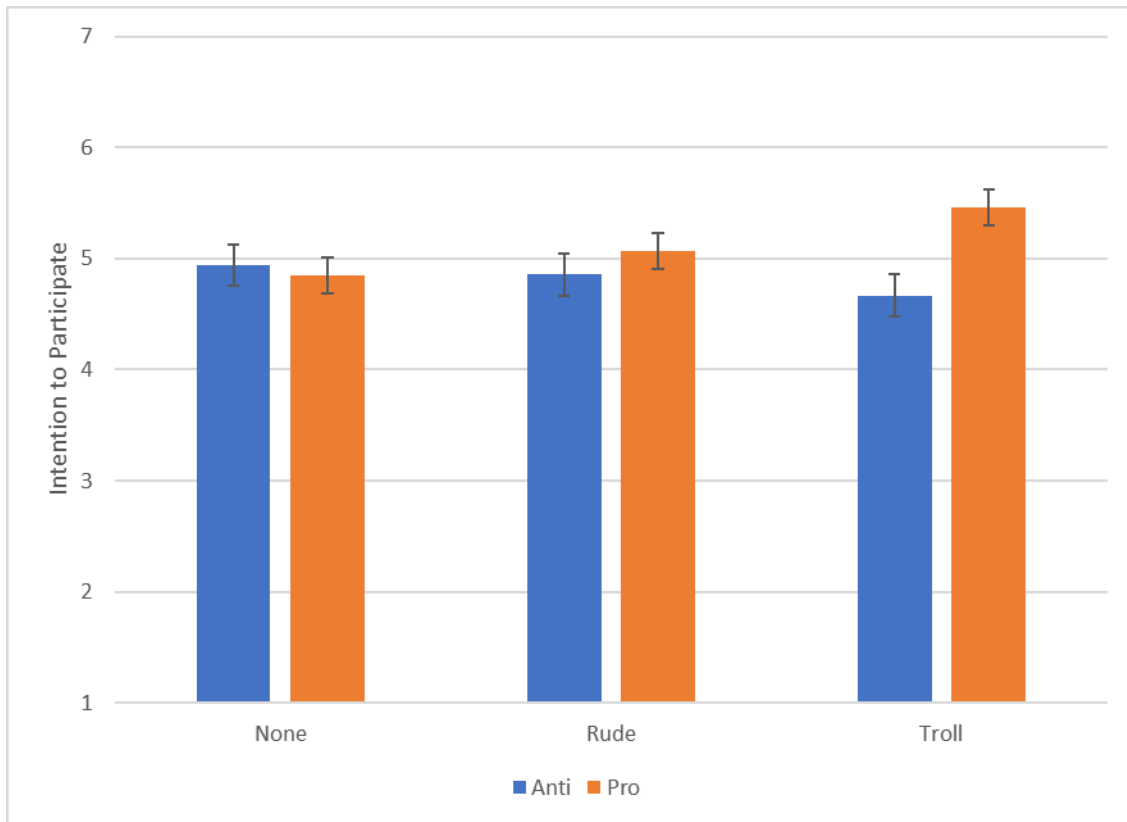


Figure 12. Ratings of participation intention by transgender identities attitude X label condition, error bars +/-1 SE.

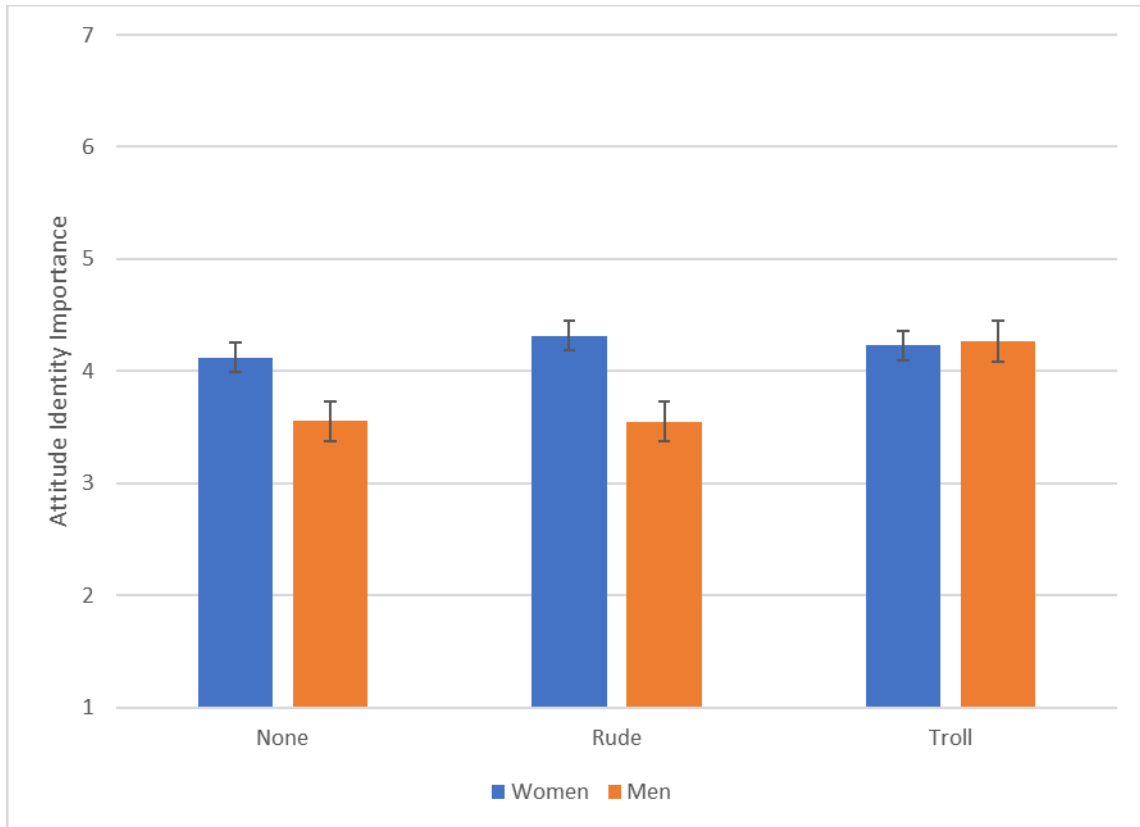


Figure 13. Ratings of pre-existing attitude identification by gender X label condition, error bars +/-1 SE.

IRB Letter



To: Dene Wamsley
From: Douglas James Adams, Chair
IRB Committee
Date: 02/13/2019
Action: **Exemption Granted**
Action Date: 02/13/2019
Protocol #: 1901171991
Study Title: Calling Out the Trolls

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.

cc: Denise R Beike, Investigator

Questionnaires

Modified SPANE-N (Diener et al., 2009)

Please think about the discussion group you have just experienced. Now indicate how much you are experiencing the following emotions, using the scale below.

Angry:	A great deal	A lot	A moderate amount	A little	None at all
Happy:	A great deal	A lot	A moderate amount	A little	None at all
Sad:	A great deal	A lot	A moderate amount	A little	None at all
Anxious:	A great deal	A lot	A moderate amount	A little	None at all

Attitude Certainty & Open-Mindedness Scale (Borah, 2014)

Please read and indicate your agreement with the following statements (1 – *strongly disagree*; 7 – *strongly agree*):

1. I feel more open to the arguments on both sides of the issue after participating in this group.
2. I got a better understanding of the issue by participating in this group.
3. I feel my opinions on this issue became stronger after participating in this group.
4. I feel more confident in my own opinion on this issue after participating in this group.

Perception of Deliberative Potential of the Discussion Scale (Hwang, Kim, & Huh, 2014)

Please read and indicate your agreement with the following statements (1 – *strongly disagree*; 7 – *strongly agree*):

1. I believe this group discussion would resolve conflicts among participants with differing views on the issue.
2. I believe this group discussion would be useful for participants to gain a better understanding of the issue.
3. I believe this group discussion would help participants see the issue from multiple perspectives.
4. I believe this group discussion would lead participants to be more open to the opposing views.

Collective Self Esteem Scale (Luhtanen & Crocker, 1992)

Please read and indicate your agreement with the following statements (1 – *strongly disagree*; 7 – *strongly agree*):

1. Being a University of Arkansas student has very little to do with how I feel about myself overall.
2. Being a University of Arkansas student is an important reflection of who I am.
3. Being a University of Arkansas student is unimportant to my sense of what kind of a person I am.
4. In general, being a University of Arkansas student is an important part of my self-image.

5. Being a member of this discussion group has very little to do with how I feel about myself overall.
6. Being a member of this discussion group is an important reflection of who I am.
7. Being a member of this discussion group is unimportant to my sense of what kind of a person I am.
8. In general, being a member of this discussion group is an important part of my self-image.
9. My opinion on this specific subject has very little to do with how I feel about myself overall.
10. My opinion on this specific subject is an important reflection of who I am.
11. My opinion on this specific subject is unimportant to my sense of what kind of a person I am.
12. My opinion on this specific subject is an important part of my self-image.

Script

1. Boot up the laboratory computer and open Google Chrome. Log onto SONA to confirm that your session is taken. Familiarize yourself with the name of your upcoming participant. Close this window completely until after your session is complete.
2. Grab a blank Informed Consent form from the box under the window. Place this in front of the computer keyboard with a pen.
3. Open the Word document in the middle of the desktop titled “ONLINE DISCUSSION SURVEY”. Copy and paste the link into a Google Chrome browser window. (The survey is also bookmarked on Google Chrome if you’d like to find it that way.) Leave the Qualtrics survey alone until the participant comes in. Close the Word document. (Here’s the link if you need it: https://uark.qualtrics.com/jfe/form/SV_5c10IVPB4WDPM3P)
4. Retrieve your participant from the lobby by calling out the first and last name of the person signed up for your research session on SONA. Double-check with the participant that they are here for the Online Discussion Study. Bring them back to the lab room you are situated in and have them sit in front of the computer.

5. Read this aloud to the participant:

“Before we begin, I need to ask you to please silence your cell phone and place it in this basket for the remainder of the study. This is to ensure that your phone does not distract you.” Once they agree, place the basket somewhere out of arm’s reach of the participant. Thank them for their cooperation.

6. Orient them to the Informed Consent form on the keyboard. Read this aloud:

“Next, the form in front of you explains your rights and risks to participating in this research today. In this experiment, you will be asked to beta-test a new chat forum. The goal of this new chat forum is to foster an environment to talk about thought-provoking topics with fellow University of Arkansas students. It’s still in early development, so we are asking for your feedback today. In this chat, you will be randomly matched with other people and asked to discuss a randomly-selected topic. This discussion will be entirely anonymous. No indicating markers about your identity will be recorded. After the discussion, you will be asked to respond to a short survey questionnaire. While there are no risks to participating in this research that greatly differ from what people experience in everyday life, the topic that is randomly selected for you today may be slightly controversial in nature. You have the right to end your participation at any moment by alerting me and letting me know you would like to stop. Do you have any questions?”

Answer any questions they may have (without revealing anything).

Continue reading:

“Please read the Informed Consent in front of you. If you agree to continue participating, please sign at the bottom. Today’s date is _____.”

7. After they have agreed and signed the Informed Consent, place it in the box under the window in the folder with other signed Informed Consents.

8. Continue reading:

“Once I leave, you can begin the study on the computer in front of you. Please read the instructions fully and carefully. When you are placed into the discussion, the browser will update every time a member writes a comment. You may need to scroll down to view each new message. You will have to wait on your fellow discussion members to write a

response during this portion, so don't be alarmed if there is a brief waiting period between comments. If you have any questions during the study, I will be in the hallway. Do you have any questions before you begin?"

Answer any questions they may have (without revealing anything). Leave the room and pull the door until it is almost closed (leave it cracked open). Wait for the participant in the hallway.

9. When the participant alerts you that they have reached the end, enter the room, close the door, and read the following aloud:

"Thank you so much for helping with this research today. As you read in your debriefing message, the discussion members you interacted with were not actually real, and the comments that were presented were made by a computer program. None of the comments you read today reflect the researchers' opinions on this given topic. PLEASE DO NOT SHARE ANY INFORMATION ABOUT THIS STUDY TO ANY OTHER INDIVIDUAL. Are there any remaining questions I can help answer today?"

Help to answer any further questions the participant may have. Once finished, thank them again for their participation and let them know they may leave.

10. Once they've left, assign the participant credit in the SONA system.