The Effects of Quote Retweet on Subsequent Posting Behavior and Morality Expression on Social Media

Yan Wu
Nanjing University
waynefire98@gmail.com

Qianzhou Du
Nanjing University
qianzhou@nju.edu.cn

Xiaohui Zhang
Arizona State University
xiaohuizhang@asu.edu

Zhongju Zhang
Arizona State University
zhongju.zhang@asu.edu

Abstract

Guided by the Threshold model and Self-justification theory, we propose and test a research model regarding the impact of online discussion activity on users' behaviors on social media. Specifically, we examine the effects of quote retweeting a tweet related to immigration policies and border issues on users' subsequent posting behaviors and morality expression. In addition, we test the moderating effect of individual threshold level and behavior-opinion inconsistency on the main effect. Results indicate that individuals, who quote retweeted the selected-topic tweets, are likely to post more topic-related tweets and express more on morality. This impact can be strengthened when individuals have higher threshold levels or larger behavior-opinion inconsistency. These findings provide both theoretical and practical implications for social media governance.

Keywords: threshold model, self-justification theory, morality expression, social media

1. Introduction

During the past decade, social media, such as Twitter, Facebook, and TikTok, has become an indispensable part of social life. Increasing numbers of people distribute and share their information, opinions, stories, and emotion online. The development of social media has brought many benefits to both firms and individuals. For example, firms can promote their own company and products through social media (Bharadwaj, El Sawy, Pavlou, & Venkatraman, 2013; Culnan, McHugh, & Zubillaga, 2010; Goh, Heng, & Lin, 2013), and they also can avoid adverse events (Abbasi, Li, Adjeroh, Abate, & Zheng, 2019) and even significantly reduce the negative effects of adverse events or negative reviews using manager's strategical response (Ravichandran & Deng, 2022). On the other hand, individuals who have a large number of followers or fans can make profits through advertising (Leung, Gu, & Palmatier, 2022; Voorveld, Van Noort, Muntinga, & Bronner, 2018) or tipping (Lin, Yao, & Chen, 2021; Lu, Yao, Chen, & Grewal, 2021).

Meanwhile, social media can backfire on society and result in terrible consequences, which can be really severe to a public extent. The spread of rumors (Oh, Agrawal, & Rao, 2013) and disinformation (Kim & Dennis, 2019; Zhang, Du, & Zhang, 2022) may worsen social crises or hurt the financial market. The recommendation algorithm and individuals' cognitive bias make it much more difficult for individuals to receive opponent opinions, and individuals are getting more extreme in their 'echo chamber' (Kitchens, Johnson, & Gray, 2020). These consequences are magnified in political fields where people tend to be more emotional and intolerant (Mutz, 2001).

Two social movements, #Black Lives Matter (BLM) and #Stop the Steal (STS), have provided practical annotations for the consequences of social media's backfire in politics. In the beginning, through tweets containing these two hashtags, participants show personal morality positions and express reasonable demands for justice. However, interactions among individuals and groups rapidly deteriorate when some of them perceive unbearable violations or conflicts against their moral norms. Meanwhile, as time goes by, each polarized group expands with the participation of users holding more extreme moral positions, which increases the amount of content filled with misleading information, hate, and harm. Participants' morality expressions correspondingly change to be more extreme or intense. Adding to the problem, they may be urged or voluntary to join extreme offline movements which finally develop into a social crisis and cause terrible consequences such as the Capitol attack and personal injuries.

Those participants in the two movements did unjust and even illegal things with the original intention of promoting morality. Without any doubt, morality is the keystone for a stable and prosperous society, in which we need to maintain fairness, moderation, and equality of rights and resources. The evolution of an individual's positing decision and morality is an inner process. It is not only influenced by outside-in information exposures but also shaped by inside-out self-reactions. Thus, it is urgent to

URI: https://hdl.handle.net/10125/103094
978-0-9981331-6-4

understand the underlying mechanism of posting behavior and morality change on social media. However, existing literature fails to answer this question well. In this study, we aim to investigate the influence of the inside-out reactions on social media users' posting behaviors and morality expressions.

In this study, we develop our research model and hypotheses based on the Threshold model (Granovetter, 1978; J. Sakoda, 1949; J. M. Sakoda, 1971) and the Self-justification theory. Specifically, the Threshold model is used to model the relationship between collective information and individual's motives and activities. This theory (González-Bailón, Borge-Holthoefer, Rivero, & Moreno, 2011; Granovetter, 1978; Schelling, 1971) argues that people are perceived as having a "threshold of importance" that they must pass before conducting a particular activity, or expressing an attitude/opinion (e.g., morality expression to some political topics). In the context of social media, passing the threshold or not is explicitly expressed as posting/quoting a tweet or not. On the other hand, it also implies a user's inner attitude to some phenomena or events, which can influence the user's subsequent activity (e.g., tweet) and morality (e.g., justice, fairness, etc.) (González-Bailón et al., 2011). This work narrows down specific research question to: **(1) how do quote retweets (threshold exceeded) influence a user's posting behavior and morality expression?** where quote retweet is treated as a strong attitude to another tweet.

Furthermore, the Threshold model suggests that individuals have different threshold levels, which can reflect the threshold intensity and result in different-level reactions (González-Bailón et al., 2011; Granovetter, 1978). The level of users' threshold indicates how important the posted content is for them or how strong users' willingness is to present their attitude. Accordingly, we raise the second research question: **(2) how does individual threshold level moderate the impacts of quote retweets on user's posting behavior and morality?** In addition, according to the Self-justification theory, if a person encounters cognitive dissonance, or a situation in which a person's behavior is inconsistent with their beliefs or opinions (e.g., political affiliation and morality), that person tends to spend more effort to justify the behavior. Therefore, the last question we want to answer is: **(3) how does the behavior-opinion inconsistency moderate the impacts of quote retweets on user's posting behavior and morality?**

To answer the three research questions, we collected data from Twitter. In detail, we collected the tweets related to a particular topic from a popular politician on border issues and immigration policy. Then, we collected the historical data of whom (treatment group) has quoted the selected-topic tweets. Meanwhile, using propensity score matching (PSM) method, we matched a control group in which persons have similar activity patterns with the persons in the treatment group but never quoted selected-topic tweets, and their historical data was collected too.

The analysis yields several interesting findings. We find that a threshold exceeding experience on a certain topic can positively influence the user's post quantity and morality intensity on the concerned topic. Additionally, users with higher threshold levels exhibit stronger changes in their behavioral patterns and moral opinions. Similarly, when the behavior-opinion inconsistency is higher, users exert more efforts (i.e., stronger behavior/attitude changes) to justify their opinions on the topic. Our findings offer a set of theoretical and practical implications. First, while previous research on social media mainly focuses on external factors' influences, we switch the gear to examine the influence of internal reactive expressions. Also, we are among the first few studies to focus on people's morality expressions on social media. Our theoretical framework combines the threshold model and the self-justification theory to help understand the underlying mechanisms of the effects. Practically, our findings help social media platforms monitor and moderate content.

2. Related Literature and Hypothesis

2.1 The Threshold Model

The decision to join a social media discussion is based on an individual's intrinsic attitude toward the discussion topic, while also influenced by the individual's dynamic interactions with other members. Thus we leverage the Threshold model (Granovetter, 1978; J. M. Sakoda, 1971) to describe this decision process. The Threshold model is usually used to model the attitudes and behaviors of individuals or groups, ranging from animal herds to human society. The classical threshold model was first introduced by Sakoda (1949), who proposed the checkerboard model which is a computational model to quantify the procedure of social interactions. In this model, each individual's attitude or behavior can be significantly changed after exceeding a certain threshold value.

Sakoda (1971) and Schelling (1971) further enrich the literature on Threshold models. Specifically, Schelling (1971) points out that the effects of individuals' interactions on their subsequent behaviors are dynamic. Following Schelling's work, Granovetter (1978) argues that people are perceived as having a threshold that they must pass before making a decision to join a collective behavior.

The threshold model can be applied in many contexts of collective behavior, including riot, residential segregation, and the spiral of silence (Granovetter, 1978). In offline contexts, this theory has been applied in situations such as the diffusion of goods (Lopez-Merino & Rouchier, 2022) and evacuation behavior in a natural disaster (Kuhlman, Marathe, Vullikanti, Halim, & Mozumder, 2022). In online social media, the threshold model theory shows more rich applications, such as information adoption (de Oliveira, Marques-Neto, & Karsai, 2022), information diffusion in social network services (Li et al., 2023) and information access inequalities (Diaz-Diaz, San Miguel, & Meloni, 2022).

2.2 Individual's Threshold Level

One important assumption of threshold model theory is that each individual has a personally-different threshold level he/she must pass before conducting certain collective behaviors, and those who did not pass the thresholds will not conduct such behaviors (Granovetter, 1978). The threshold level is defined as the proportion of participants an individual needs to see before he/she finally decides to conduct the same behavior. For each individual, the threshold level is a kind of 'disposition' which is stable and keeps on a certain level long enough for the collective behavior or social events to reach an equilibrium (Granovetter, 1978).

Although each individual's threshold level shows its stability, individuals who have decided to join a social event are still allowed to show great behavioral and belief changes. In an example of the decision to join a riot, Granovetter (1978) points out the possibility for an individual of a lynch mob to have great changes in their values, preferences and behaviors, because they are only bringing 'contingent dispositions' into the certain situations. Therefore, their individual threshold levels still show continuity during the social event, while preferences and values can be changed. The possibility of great behavior and belief changes is also supported in previous research, which finds that when individuals exceed thresholds, they might be more likely to participate in these social events actively (Granovetter, 1978; Granovetter & Soong, 1983).

Another important characteristic of individual threshold level is that it is individually different. Firstly, different demographic factors, including social class, education, occupation, and social position, may affect individual threshold levels (Granovetter, 1978). These factors may lead to great variation in individuals' thresholds and correspondingly lead to different outcomes. Meanwhile, individual's threshold level is also context based, which means the same individual may hold different thresholds facing different situations. Another important determinant is the individually hidden process of comparing the benefits and costs to form the threshold, which indicates that two individuals with the same threshold may hold different benefits and costs in joining the social event. These individually different characteristics indicate that individuals with different threshold levels may experience different behavior and belief changes after joining a social event.

2.3 Self-justification Theory

The behavior of joining a social media discussion regards another important aspect where individual's behaviors and beliefs will influence and shape each other. Therefore, we leverage the Self-justification theory to discuss the circumstances where an individual experiences greater behavior and belief changes after joining a social discussion.

According to the theory of cognitive dissonance (Festinger, 1957), cognitive dissonance is the perception of information conflict, in which the information includes individual's opinions, ideas, thoughts, beliefs, standpoints, values, etc. When a person's behavior conflicts with one or more of those things, that person might experience psychological stress and discomfort which is triggered by the clashing between person's beliefs or values and the information he/she perceives. To reduce this discomfort, individuals will do all in their power to resolve the contradiction between behavior and opinions. The Self-justification theory (Staw & Fox, 1977) suggests that individuals tend to self-justify their behaviors, when inconsistency exists between their behaviors and beliefs. The greater inconsistency exists, the more efforts individuals will spend decreasing cognitive dissonance.

Self-justification theory has been tested in a lot of experiments across species (Lieberman, Ochsner, Gilbert, & Schacter, 2001). It receives more attention in the context of social media, where mass contents are created and users are facing large amounts of conflicting opinions every day. For example, it provides an explanation for users' preference for selective information exposure (Jeong, Zo, Lee, & Ceran, 2019), which is a commonly used mechanism to explain the 'echo chamber' effect that describes people's refusion against multiple information sources in political discussions (Colleoni, Rozza, & Arvidsson, 2014; Guo, A. Rohde, & Wu, 2020; Himelboim, McCreery, & Smith, 2013). These studies have illustrated the effect of cognitive dissonance in the

context of information conflicts on social media, centering on the information consumption process.

However, there is little prior research exploring the effect of conducting collective behavior, such as joining a social media discussion, from the perspective of Self-justification theory. To be more specific, the behavior of joining a social media discussion may influence individual's subsequent beliefs and behaviors. As is illustrated in many experiments, people tend to hold an increased preference over time for the choices they have made (Brehm, 1956; Festinger, 1957). This tendency of individuals' shift in their preferences to align with actions has also been explained with a rational choice approach in a political context (Acharya, Blackwell, & Sen, 2018).

These studies indicate that the joining behavior itself may cause differences in subsequent behaviors and beliefs between users who have joined the social media discussion and those who have not. And the intensity of cognitive dissonance may amplify such differences. By incorporating the Self-justification theory with the Threshold model theory, this study sheds light on the impact of joining social media discussion after individuals pass their thresholds.

2.4 Research Model and Hypothesis

In a social media context, we base our study on the Threshold model theory and Self-justification theory to explain individual's decision of joining a discussion, as well as its impact on subsequent morality-related behaviors and beliefs. In this study, we choose the context of fairness, and set the specific interested topic as border issues and immigration policies (referred to as 'selected topic'), a topic which is frequently discussed by a popular politician (i.e., Donald Trump and Joe Biden) and his followers throughout his presidency. Thus, individual's decision of joining the discussion of the selected topic is viewed as a process of 'collective behavior'. If an individual has passed his/her threshold, he/she will join the discussion of the selected topic.

It needs to be clarified on how to observe whether individual has passed the threshold. To join a discussion means to express attitudes explicitly, including giving out likes, retweeting the supported tweets, or quote retweeting tweets. Among all these behaviors, quote retweet is viewed as an obvious participation into the discussion because it contains both the behavior participation (to retweet) and morality expressions (individual's comments). Therefore, in the context of social media, we use this confirmed activity (e.g., quote retweet) as an indicator to reflect whether that individual exceeds his/her threshold or not. If a user quote retweets a selected topic tweet from the popular politician, we think that the user has passed the threshold, thus has made the decision to join the discussion.

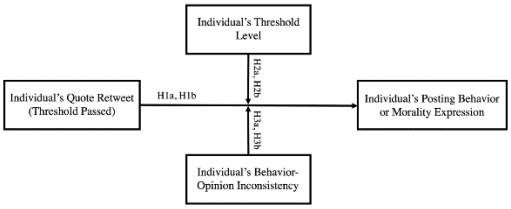

Figure 1. Research Model

With all above clear, we propose our research model as what Figure 1 shows. Firstly, as is illustrated in the literature review, individuals must pass a certain threshold to finally decide to join a discussion. As is predicted by the Threshold model, individuals will be more active as they have joined the discussion (Granovetter, 1978; Granovetter & Soong, 1983). From another aspect, the Self-justification theory explains that this change may come from individual's effort to justify his/her decision to join the discussion, such as shifting preferences (Acharya et al., 2018) or changing behaviors to be more aligned with the decision of joining. In a social media context, as individuals exceed their thresholds and conduct the behavior of joining the discussion (quote retweet behavior), they will show more stress on morality in the discussions of the quoted topic (i.e., the selected topic) in the following days. Meanwhile, they will increase their frequency of relevant discussions, such as posting more tweets about the quoted topic. Therefore, we hypothesize:

H1a: Individual's **threshold-exceeding** (quote retweet behavior) is associated with **more tweets about** the quoted topic in a subsequent time interval (posting behavior change).

H1b: Individual's **threshold-exceeding** (quote retweet behavior) is associated with **more morality expression** on the quoted topic in a subsequent time interval (morality expression change).

As is discussed previously, threshold levels show differences across individuals, affected by demographic factors and situations. Studies have shown that as individuals enter the collective behavior in different thresholds, they may act differently. For example, González-Bailón et al. (2011) apply the Threshold model to quantify individuals' engagement in social movements, such as the mass mobilization in Moldova in 2009, the Arabspring, and the color revolutions across western countries. They statistically show that individuals indeed have their own unique threshold levels to involve a particular social movement. Furthermore, they find that individuals with higher threshold levels engage to mass mobilization more actively than the ones with lower

threshold levels after exceeding their threshold, even though they react later than the ones with lower threshold levels. Similar to the social movement context, individuals who join the discussion on social media with higher thresholds may show greater increase in both posting behaviors and stress on morality in their subsequent discussions.

Following the above discussion, we hypothesize:

H2a: Individual's **threshold level** positively moderates the relationship between quote retweet behavior and **posting behavior change** on the quoted topic. Individuals with higher threshold levels have higher increase in subsequent tweets about the quoted topic.

H2b: Individual's **threshold level** positively moderates the relationship between quote retweet behavior and **morality expression change** on the quoted topic. Individuals with higher threshold levels have higher increase in their stress on morality in subsequent discussions.

As is discussed in the literature of the Self-justification theory, individuals have a tendency to shift their preferences to be aligned with actions to decrease cognitive dissonance (Acharya et al., 2018). The greater inconsistency between behaviors and opinions exists, the more efforts individuals will spend on decreasing the inconsistency. In our research context, taking quote retweet behavior as a pre-determined action which has been made by the individual, behavior-opinion inconsistency is triggered when individuals have decided to join the discussion led by the popular politician while holds a quite different morality opinions comparing to that politician. Specifically, the quote retweet behavior consists of two parts, including the quoted content written by the popular politician and the comment content written by a particular user. We measure individual's behavior-opinion inconsistency as the morality opinion difference between the quoted content and individual's comment. The higher inconsistency an individual has, the more efforts the individual will make to conquest it, which will show in their subsequent behaviors and beliefs. Therefore, we assume that individual's behavior-opinion inconsistency strengthens the impact of exceeding thresholds on individual's posting behaviors and morality expressions on the selected-topic tweets. We hypothesize that:

H3a: Individual's **behavior-opinion inconsistency** positively moderates the relationship between quote retweet behavior and **posting behavior change** on the quoted topic. Individuals with higher behavior-opinion inconsistency have higher increase in their subsequent tweets about the quoted topic.

H3b: Individual's **behavior-opinion inconsistency** positively moderates the relationship between quote retweet behavior and **morality expression change** on the quoted topic. Individuals with higher behavior-opinion inconsistency have higher increase in their stress on morality in their subsequent discussions.

3. Research Context and Data

3.1. Data Collection and Process

To test our research hypotheses, we collect data from Twitter, which is one of the largest microblogging and social networking platforms in the World, and on which users post and interact with messages known as "tweets". By the start of 2019, Twitter had more than 330 million monthly active users. On Twitter, each user can express and share information and opinions related to politics, economy, business, sports, entertainment, etc. Twitter users can post, reply, retweet (without a comment), and quote retweet (with a comment) a tweet. In this study, we first collected all tweets from a very popular politician's twitter account from May 2009 to January 2020. During this period, he posted 56,571 tweets to the forum, which cover several important topics, such as trade war, border issues and immigration policy, pandemic policy, etc.

In this study, we filter all tweets of the selected politician *X* using the four key words, including "immigration", "border", "border security", and "crime", and then we get the 500 most related tweets about border issues and immigration policy. We collect 2,684 *X*'s followers who quote retweet the border-related tweets as the treatment group, and we apply the propensity score matching (PSM) method to generating a comparison group of users who are also *X*'s followers, but they never quote retweet any *X*'s tweet during our study period. The matching covariates include user tenure, number of followers, number of followees, and number of statuses. Then, we successfully match 1,749 users and 3,264 users in the treatment group and control group, respectively. Finally, we crawl the historical tweets data for these users and get 3,522,450 tweets in our study period. The procedure of data collection and pre-processing are shown in Figure 2.

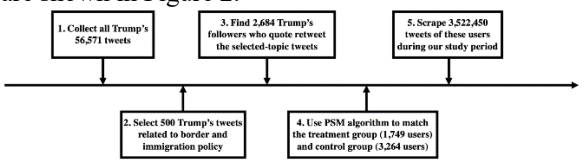

Figure 2. The Procedure of Data Collection and Pre-processing

3.2. Variable Measures

3.2.1. Dependent Variables To evaluate the impacts of exceeding thresholds, we measure all variables based on the collected Twitter data. First of all, we have two dependent variables, including individual's posting behavior change and individual's morality expression change. The dependent variable, individual's posting behavior change, is measured by the change of the percentage of individual's selected-topic (i.e., border issues and immigration policy) tweets before and after the day a user quote retweets one of the selected-topic tweets. As shown in Figure 3, setting the quote retweet day as the cut-off point, we divide a user's whole observation time into two periods: "before period", a 180-day period from 180 days to 1 day prior to user's quote retweet day, and "after period", a 4-day period from the quote retweet day to 3 days after. Then, posting behavior change is defined as the difference of the selected-topic related posting behaviors between "after period" and "before period", using the following formula:

$$behavior\ change_i = \frac{1}{4} * \sum_{k=t}^{k=t+3} \frac{Num\_of\_Tweets_{i,k,s}}{Num\_of\_Tweets_{i,k,all}} - \frac{1}{180} * \sum_{k=t-180}^{k=t-1} \frac{Num\_of\_Tweets_{i,k,s}}{Num\_of\_Tweets_{i,k,all}} \quad (1)$$

where i denotes a particular $user_i$; t indicates the quote retweet day and k is the date; $Num\_of\_Tweets_{i,k,s}$ is the number of the selected-topic tweets posted by $user_i$ on day_k ; $Num\_of\_Tweets_{i,k,all}$ denotes the number of all tweets posted by $user_i$ on day_k.

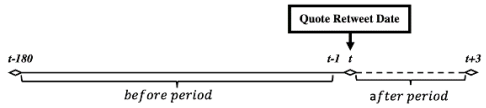

Figure 3. The Illustration of Time Periods

The second dependent variable is morality expression change, referring to the difference of user's morality expressions between "before period" and "after period". Morality expression means the extent to which a user emphasizes on the prescriptive aspect of morality and call for morality. Higher morality expression shows user's higher-level appealing or extremeness in their morality standpoint. Using a dictionary-based natural language processing (NLP) software LIWC, we measure each tweet's morality expression. Specifically, the morality wordlist is from Moral Foundation Dictionaries 2.0 (Frimer, 2019) which is based on Moral Foundation Theory (Graham, Haidt, & Nosek, 2009). This dictionary-based approach measures a given tweet on five bipolar (virtue or vice) dimensions of human morality, including virtue (e.g., care, fairness, loyalty, authority, and purity) and vice (harm, cheating, betrayal, subversion, and degradation). As discussed before, in the context of immigration policies and border issues, we select the dimension of fairness/cheating as the morality expression measurement. Then, the moral expression $morality_f$ of a given $Tweet_f$ is defined as the following formula:

$$morality_f = \frac{morality\_virtue_f}{morality\_virtue_f + morality\_vice_f} \quad (2)$$

where $morality\_virtue_f$ and $morality\_vice_f$ denote the virtue and vice scores on the dimension of fairness/cheating of $Tweet_f$, respectively. $morality_f$ is a continuous variable between 0 and 1 and represents the degree of morality expressions in a tweet. A higher $morality_f$ indicates a tweet shows higher or more extreme appealing for morality, and a lower Image means that the tweet is weaker in its call for morality. User's daily morality expression is the average of his/her selected-topic tweets' $morality$. Finally, morality expression change is calculated as the difference of morality expression between the two periods, using the following formula:

$$morality\ change_i = \frac{1}{4} * \sum_{k=t}^{k=t+3} \sum_{f \in Tweets_{i,k,s}} \frac{morality_f}{Num\_of\_Tweets_{i,k,s}} - \frac{1}{180} * \sum_{k=t-180}^{k=t-1} \sum_{f \in Tweets_{i,k,s}} \frac{morality_f}{Num\_of\_Tweets_{i,k,s}} \quad (3)$$

where i denotes a particular $user_i$; t indicates the quote retweet day and k is the date; $morality_f$ denotes morality expression of $Tweet_f$; $Tweets_{i,k,s}$ are the selected-topic tweets posted by $user_i$ on day_k.

3.2.2. Independent and Control Variables Firstly, **Treat** refers to whether an individual passed the threshold (e.g., he/she quote retweets one of the selected-topic tweets posted by a popular politician) or not. If yes, that user is considered as a member of the treatment group. For the control group, we apply propensity score matching (PSM) method to generating a comparison group of users who are also the popular politician's followers but have never quote retweeted any of his tweets during our study period. The matching covariates include user tenure, number of followers, number of followees, number of statuses. In this study, we measure the variable, **Treat**, using a dummy variable to indicate whether a user exceeds a threshold or not. One represents the users have exceeded their thresholds, and zero otherwise.

Threshold level refers to the threshold value that a user needs to exceed before quote retweeting a tweet. With a higher threshold level, users need to see more people who have decided to join the discussion to conduct the same collective behavior (i.e. to quote retweet a selected-topic tweet). Therefore, to measure individual's threshold level, we choose the total number of quotes received by the selected-topic tweets posted by the popular politician during the 10-day

interval before the individual's quote retweet day. The threshold level is calculated as the following formula:

$$threshold\ level_i = \sum_{k=t-10}^{k=t-1} Num\_of\_Quotes_{i,k,s} \quad (4)$$

where i denotes a particular $user_i$; t indicates the quote retweet day and k is the date; $Num\_of\_Quotes_{i,k,s}$ is the total number of quotes received by the selected-topic tweets of the popular politician on day_k.

Behavior-opinion inconsistency refers to the inconsistency triggered by a user's behavior of joining the discussion and his/her opinion difference with the popular politician. As the individual has decided to join the discussion (passed the threshold), higher opinion difference can trigger higher level of cognitive dissonance. Thus, we measure behavior-opinion inconsistency by calculating the absolute percentage difference of morality expressions between the quoted content written by the popular politician and the comment written by that individual, as the following formula:

$$behavior\ opinion\ inconsistency_i = \frac{\left|morality_{comment_i} - morality_{quoted\_content_i}\right|}{morality_{quoted\_content_i}} \quad (5)$$

where i denotes a certain $user_i$; $morality_{comment_i}$ indicates the morality expression of the comment part; $morality_{quoted\_content_i}$ denotes the morality expression of the quoted retweet.

For control variables, we first control user's **political affiliation**, which refers to user's political preference towards a certain political party. To identify user's political affiliation, we apply the classic deep learning method Long Short-Term Memory (LSTM) (Hochreiter & Schmidhuber, 1997; Huang, Shen, & Deng, 2019) to implementing a text classification model. Specifically, we collect the training dataset that consists of 12,000 archived tweets with the presidential election slogans (we choose one slogan #MAGA for the Republican party, and three slogans including #nomarkely, #ourbestdaysstilllieahead and #battleforthesoulofthenation for the Democratic party) in 2016 and 2020. Each party has 6,000 tweets with a slogan. Based on the 5-fold cross-validation setting, the trained model performs effectively in identifying twitter users' politic affiliation (e.g., accuracy=0.73 and f1-score=0.74). For each tweet, we estimate the political affiliation probability using the LSTM model, in which 1 represents the probability that a tweet inclines to Democratic party and 0 means that the tweet inclines to the Republican party. Finally, a user's general $political\ affiliation_i$ is calculated as the mean political affiliation value of the archived tweets sample, using the following formula:

$$political\ affiliation_i = \frac{\sum_{f \in sampled\_archived\_tweets_i} political\_affiliation\_probability_f}{1000} \quad (6)$$

where i denotes a certain $user_i$; $sampled\_archived\_tweets_i$ is a 1000-tweet set randomly sampled from the $user_i$'s archived tweets; $political\_affiliation\_probability_f$ indicates the probability of $Tweet_f$. Second, we control user's **tenure**, which refers to how long a user joined on Twitter community. It is defined as the number of days from a user's twitter registration to that user's quote retweet day. Third, user's **age** and **gender** are under control. User's age and gender are calculated by following the instruction of previous work(Wang et al., 2019). Gender is a binary variable (e.g., 0 and 1 represent male and female, respectively.), and age is a categorical variable (e.g., 1, 2, 3, and 4 denotes the age equal to or smaller than 18, between 19 and 29, between 30 and 39, and equal to or larger than 40, respectively). In addition, we also control the **friend** number and **likes** number. Finally, we control for **time fixed effect** on a year-month level to capture different trends across the time. We summarize the statistics of key variables in our model in Table 1. Our final dataset consists of 5013 users, 34.9% of which belong to the treated group.

Table 1. Descriptive statistics of key variables

	Mean(SD)	Min.	Max.
behavior change	0.007(0.035)	-0.197	0.327
morality change	0.002(0.025)	-0.062	0.5
treat	0.349(0.477)	0	1
threshold level	23.079(77.35)	0	877
behavior-opinion inconsistency	0.041(0.163)	0	1
political affiliation	0.437(0.094)	0	1
tenure	2093.599(1071.932)	0	4669
likes	19364.21(42939.92)	0	727591
friends	1421.787(4951.109)	0	231275
age	2.881(1.092)	1	4
gender	0.333(0.471)	0	1

4. Empirical Analysis

4.1. Main Results

To test the proposed hypotheses, we conduct an empirical analysis to estimate the impacts of exceeding thresholds (e.g., quote retweet) on individual's subsequent behavior and belief changes, including posting behavior change and morality expression change. We apply an OLS model

containing a dummy (treat) indicating whether the user belongs to the treated group (e.g., 1 is treat and 0 is control). The "treat" term thus can estimate the treatment effect (the difference on group levels). Intuitively, this term captures the real differences between those who passed the threshold level and those who did not. Our econometric specification is as follows:

$$E(behavior\ change_i|x_i) = \beta_1 treat_i + \beta_2 threshold\ level_i + \beta_3 behavior\ opinion\ inconsistency_i + \beta_4 politic\ affiliation_i + \beta_5 tenure_i + \beta_6 likes_i + \beta_7 friends_i + \beta_8 age_i + \beta_9 gender_i + \beta_{10} treated * threshold\ level_i + \beta_{11} treated * behavior\ opinion\ inconsistency_i + \alpha_i + \varepsilon_i \quad (7)$$

$$E(morality\ change_i|x_i) = \beta_1 treat_i + \beta_2 threshold\ level_i + \beta_3 behavior\ opinion\ inconsistency_i + \beta_4 politic\ affiliation_i + \beta_5 tenure_i + \beta_6 likes_i + \beta_7 friends_i + \beta_8 age_i + \beta_9 gender_i + \beta_{10} treated * threshold\ level_i + \beta_{11} treated * behavior\ opinion\ inconsistency_i + \alpha_i + \varepsilon_i \quad (8)$$

(Note: i denotes the $user_i$; α_i denotes the year-month dummy of user's quote retweet day)

In our econometric specification, β_1 captures the treatment effect of exceeding thresholds, β_{10} captures the moderation effect of threshold level on the treatment effect, β_{11} captures the moderation effect of behavior-opinion inconsistency on the treatment effect. Robust standard errors are used in model estimation.

Table 2. Results of Empirical Analysis

	Model 1	Model 2	Model 3
DV	behavior change	morality change in the selected-topic tweets	morality change in all tweets
treat	0.443 (0.079)***	0.170 (0.045)***	0.197 (0.033)***
threshold level	-0.147 (0.073)**	-0.050 (0.036)	-0.052 (0.031)
behavior-opinion inconsistency	-0.034 (0.017)*	0.001 (0.005)	0.002 (0.011)
treat*threshold level	0.133 (0.028)***	0.116 (0.040)***	0.065 (0.029)**
treat*behavior-opinion inconsistency	0.176 (0.045)***	0.150 (0.094)+	0.306 (0.093)***
control variables			
politic affiliation	-0.024 (0.015)	-0.000 (0.013)	0.004 (0.008)
tenure	-0.054 (0.032)*	-0.035 (0.026)	-0.040 (0.017)**
likes	-0.034 (0.015)**	-0.011 (0.020)	-0.022 (0.012)*
friends	-0.001 (0.016)	0.005 (0.015)	-0.013 (0.018)
age	0.015 (0.010)	0.009 (0.015)	0.027 (0.013)**
gender	0.004 (0.028)	0.016 (0.021)	-0.002 (0.023)
R2	0.103	0.034	0.064
Adj. R2	0.092	0.022	0.052
nobs	5013	5013	5013
Notes: ***p< 0.01, **p<0.05, *p<0.1, +p<0.15			

Main results are reported with Model 1 and Model 2 in Table 2. Main results confirm the significance of positive treatment effects in both behavior change and morality change (Model 1: β_1=0.443, p<0.01; Model 2: β_1=0.170, p<0.01). Thus, H1a and H1b is supported.

The moderation effect of threshold level is confirmed to be significantly positive (Model 1: β_{10} =0.133, p<0.01; Model 2: β_{10} =0.116, p<0.01). Thus, H2a and H2b is supported. The moderation effect of behavior-opinion inconsistency is positive and significant in Model 1(Model 1: β_{11} =0.176, p<0.01). Thus, H3a is supported.

However, the coefficient is not significant in Model 2(Model 2: β_{11} =0.150, p=0.115). The coefficient becomes significant when including only one interaction term of treat*behavior-opinion inconsistency in Model 2(β_{11}=0.159, p<0.1). Thus, H3b is partially supported.

4.2. Additional Results

In this subsection, we conduct additional analysis to see whether the treatment (namely exceeding thresholds) has spillover effects on morality expressions in all tweets or not. First, we define the third dependent variable, morality change in all tweets, as follows:

$$morality\ change_{i,all} = \frac{1}{4} * \sum_{k=t}^{k=t+3} \sum_{f \in Tweets_{i,k,all}} \frac{morality_f}{Num\_of\_Tweets_{i,k,all}} - \frac{1}{180} * \sum_{k=t-180}^{k=t-1} \sum_{f \in Tweets_{i,k,all}} \frac{morality_f}{Num\_of\_Tweets_{i,k,all}} \quad (9)$$

where i denotes a particular $user_i$; t indicates the quote retweet day and k is the date; $morality_f$ denotes moral expression of $Tweet_f$; $Tweets_{i,k,all}$ are all tweets posted by $user_i$ on day_k.

Then, we conduct a similar analysis by replacing $morality\ change_i$ with $morality\ change_{i,all}$. The results are summarized with Model 3 in Table 2. From the results, we can find H1b, H2b, and H3b are fully supported (Model 3: β_1=0.197, p<0.01; β_{10}=0.065, p<0.01; β_{11}=0.306, p<0.01). Although the quoted content is only about the border issues and immigration policy, the results show that, if a user passes the threshold, he/she not only appeals more for morality in tweets of the same topic, but also will increase posting behaviors and morality expressions in tweets of other topics. In addition, user's behavior-opinion inconsistency and individual threshold level also can positively influence such increases.

5. Discussions and Conclusions

This study makes efforts to investigate the mechanism that influences user's posting behavior and morality expressions on social media. Different from prior research about social media, we focus on studying what will happen or change at the subsequent time after a user makes the decision to join a social discussion. Based on the Threshold model and Self-justification theory, we construct a conceptual research model integrating exceeding thresholds, threshold levels, and behavior-opinion inconsistency to estimate a user's subsequent behaviors and beliefs.

Empirical results support our hypotheses. Firstly, exceeding thresholds can positively impact a user's posting activities and morality expression of the selected-topic tweets. Secondly, threshold levels and behavior-opinion inconsistency positively moderate such impacts. Additionally, spillover effect is tested, which means that although exceeding thresholds is just involved with discussing border issues and immigration policy, it also impacts a user's subsequent morality expressions of other topics.

Our work sheds light on the impact of social media discussion on individuals, and illustrates the circumstances where users become overreacted in morality expressions, thus possibly causing social media to backfire in the form of social crisis and other terrible consequences.

5.1. Implications for Theory

This study provides three main theoretical implications. First, to our best knowledge, this is one of the first studies researching on the impacts of exceeding thresholds on subsequent behaviors. Previous relevant studies focus primarily on exploring what external factors influence a user to perform a particular activity. This study shows empirical evidence that exceeding a threshold can change users' subsequent morality-related behaviors and expressions. Second, we leverage the Threshold model to construct the conceptual framework. We apply this theory into the context of social media by viewing quote retweet as an indicator of user's decision to join social discussions, and explore changes afterwards. Last but not least, our empirical results also confirm the effectiveness of the Self-justification theory.

5.2. Implications for Practice

The practical implications of our study mainly include two parts. Firstly, our study provides online platform regulators some insights into understanding the prevailing of a social media discussion and its impacts on individual users. Such insights may help create a healthier environment for social media discussions. For example, to prevent situations from getting worse, regulators should focus on individuals with higher thresholds and cognitive dissonance because they are more likely to be extreme.

Secondly, our study helps key opinion leaders like broadcaster, youtuber and firm official accounts understand their followers. Those who quote retweet or reply to their content tend to do more related discussions subsequently. Thus, focusing on them helps key opinion leaders foster a more durable discussion that will show benefit in the future.

5.3. Limitations and Future Work

This study also has several limitations. First, this work just considers the tweets about border issues and immigration policy. Future studies may explore the tweets on other topics, such as race and violence, pandemic policy, climate change, etc. Involving more topics can test the robustness and enrich research contexts. The second limitation is that the results only partially support H2b. More analysis should be made to understand the hidden mechanism. The third limitation comes from shallow existing findings on the spillover effect of exceeding thresholds. Cross-topic difference of spillover effect needs more analysis. Last but not least, we measure the dependent variables using only one way(e.g., from t-180 to t-1 and from t to t+3). Future research can test the hypotheses using different time intervals.

6. References

Abbasi, A., Li, J., Adjeroh, D., Abate, M., & Zheng, W. (2019). Don't Mention It? Analyzing User-Generated Content Signals for Early Adverse Event Warnings. Information Systems Research, 30, 1007–1028.

Acharya, A., Blackwell, M., & Sen, M. (2018). Explaining Preferences from Behavior: A Cognitive Dissonance Approach. The Journal of Politics, 80, 400–411.

Bharadwaj, A., El Sawy, O. A., Pavlou, P. A., & Venkatraman, N. v. (2013). Digital business strategy: Toward a next generation of insights. MIS Quarterly, 471–482.

Brehm, J. W. (1956). Postdecision changes in the desirability of alternatives. The Journal of Abnormal and Social Psychology, 52, 384–389.

Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data: Political Homophily on Twitter. Journal of Communication, 64, 317–332.

Culnan, M. J., McHugh, P. J., & Zubillaga, J. I. (2010). How large US companies can use Twitter and other social

media to gain business value. MIS Quarterly Executive, 9.
de Oliveira, J. F., Marques-Neto, H. T., & Karsai, M. (2022). Measuring the effects of repeated and diversified influence mechanism for information adoption on Twitter. Social Network Analysis and Mining, 12, 16.
Diaz-Diaz, F., San Miguel, M., & Meloni, S. (2022). Echo chambers and information transmission biases in homophilic and heterophilic networks. Scientific Reports, 12, 9350.
Festinger, L. (1957). A theory of cognitive dissonance (Vol. 2). Stanford university press.
Frimer, J. (2019). Moral Foundations Dictionary 2.0. https://doi.org/10.17605/OSF.IO/EZN37
Goh, K.-Y., Heng, C.-S., & Lin, Z. (2013). Social media brand community and consumer behavior: Quantifying the relative impact of user-and marketer-generated content. Information Systems Research, 24, 88–107.
González-Bailón, S., Borge-Holthoefer, J., Rivero, A., & Moreno, Y. (2011). The dynamics of protest recruitment through an online network. Scientific Reports, 1, 1–7.
Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. Journal of Personality and Social Psychology, 96, 1029–1046.
Granovetter, M. (1978). Threshold models of collective behavior. American Journal of Sociology, 83, 1420–1443.
Granovetter, M., & Soong, R. (1983). Threshold models of diffusion and collective behavior. Journal of Mathematical Sociology, 9, 165–179.
Guo, L., A. Rohde, J., & Wu, H. D. (2020). Who is responsible for Twitter's echo chamber problem? Evidence from 2016 U.S. election networks. Information, Communication & Society, 23, 234–251.
Himelboim, I., McCreery, S., & Smith, M. (2013). Birds of a Feather Tweet Together: Integrating Network and Content Analyses to Examine Cross-Ideology Exposure on Twitter. Journal of Computer-Mediated Communication, 18, 40–60.
Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9, 1735–1780.
Huang, T., Shen, G., & Deng, Z.-H. (2019). Leap-lstm: Enhancing long short-term memory for text categorization. ArXiv Preprint ArXiv:1905.11558.
Jeong, M., Zo, H., Lee, C. H., & Ceran, Y. (2019). Feeling displeasure from online social media postings: A study using cognitive dissonance theory. Computers in Human Behavior, 97, 231–240.
Kim, A., & Dennis, A. R. (2019). Says who? The effects of presentation format and source rating on fake news in social media. MIS Quarterly, 43, 1025–1039.
Kitchens, B., Johnson, S. L., & Gray, P. (2020). Understanding Echo Chambers and Filter Bubbles: The Impact of Social Media on Diversification and Partisan Shifts in News Consumption. MIS Quarterly, 44, 1619–1649.
Kuhlman, C. J., Marathe, A., Vullikanti, A., Halim, N., & Mozumder, P. (2022). Natural disaster evacuation modeling: The dichotomy of fear of crime and social influence. Social Network Analysis and Mining, 12, 13.
Leung, F. F., Gu, F. F., & Palmatier, R. W. (2022). Online influencer marketing. Journal of the Academy of Marketing Science, 1–26.
Li, H., Xia, C., Wang, T., Wen, S., Chen, C., & Xiang, Y. (2023). Capturing Dynamics of Information Diffusion in SNS: A Survey of Methodology and Techniques. ACM Computing Surveys, 55, 1–51.
Lieberman, M. D., Ochsner, K. N., Gilbert, D. T., & Schacter, D. L. (2001). Do Amnesics Exhibit Cognitive Dissonance Reduction? The Role of Explicit Memory and Attention in Attitude Change. Psychological Science, 12, 135–140.
Lin, Y., Yao, D., & Chen, X. (2021). Happiness begets money: Emotion and engagement in live streaming. Journal of Marketing Research, 58, 417–438.
Lopez-Merino, P., & Rouchier, J. (2022). The diffusion of goods with multiple characteristics and price premiums: An agent-based model. Applied Network Science, 7, 11.
Lu, S., Yao, D., Chen, X., & Grewal, R. (2021). Do larger audiences generate greater revenues under pay what you want? Evidence from a live streaming platform. Marketing Science, 40, 964–984.
Mutz, D. C. (2001). Facilitating Communication across Lines of Political Difference: The Role of Mass Media. American Political Science Review, 95, 97–114.
Oh, O., Agrawal, M., & Rao, H. R. (2013). Community intelligence and social media services: A rumor theoretic analysis of tweets during social crises. MIS Quarterly, 407–426.
Ravichandran, T., & Deng, C. (2022). Effects of Managerial Response to Negative Reviews on Future Review Valence and Complaints. Information Systems Research.
Sakoda, J. (1949). Minidoka: An analysis of changing patterns of social behavior (PhD Thesis). PhD thesis, University of California.
Sakoda, J. M. (1971). The checkerboard model of social interaction. The Journal of Mathematical Sociology, 1, 119–132.
Schelling, T. C. (1971). Dynamic models of segregation. Journal of Mathematical Sociology, 1, 143–186.
Staw, B. M., & Fox, F. V. (1977). Escalation: The determinants of commitment to a chosen course of action. Human Relations, 30, 431–450.
Voorveld, H. A., Van Noort, G., Muntinga, D. G., & Bronner, F. (2018). Engagement with social media and social media advertising: The differentiating role of platform type. Journal of Advertising, 47, 38–54.
Wang, Z., Hale, S., Adelani, D. I., Grabowicz, P., Hartman, T., Flöck, F., & Jurgens, D. (2019). Demographic Inference and Representative Population Estimates from Multilingual Social Media Data. The World Wide Web Conference, 2056–2067. San Francisco CA USA: ACM.
Zhang, X., Du, Q., & Zhang, Z. (2022). A Theory-Driven Machine Learning System for Financial Disinformation Detection. Production and Operations Management. In press.