

“Algorithms ruin everything”: #RIPTwitter, Folk Theories, and Resistance to Algorithmic Change in Social Media

Michael A. DeVito
Northwestern University
Evanston, IL, USA
devitom@u.northwestern.edu

Darren Gergle
Northwestern University
Evanston, IL, USA
dgergle@northwestern.edu

Jeremy Birnholtz
Northwestern University
Evanston, IL, USA
jeremyb@northwestern.edu

ABSTRACT

As algorithmically-driven content curation has become an increasingly common feature of social media platforms, user resistance to algorithmic change has become more frequent and visible. These incidents of user backlash point to larger issues such as inaccurate understandings of how algorithmic systems work as well as mismatches between designer and user intent. Using a content analysis of 102,827 tweets from #RIPTwitter, a recent hashtag-based backlash to rumors about introducing algorithmic curation to Twitter’s timeline, this study addresses the nature of user resistance in the form of the complaints being expressed, folk theories of the algorithmic system espoused by users, and how these folk theories potentially frame user reactions. We find that resistance to algorithmic change largely revolves around expectation violation, with folk theories acting as frames for reactions such that more detailed folk theories are expressed through more specific reactions to algorithmic change.

Author Keywords

Algorithms; algorithm awareness; folk theories; technology continuance; user resistance; social media; algorithmic curation; expectation violation; machine classification

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI):

INTRODUCTION

Algorithmically-driven content curation systems such as the Facebook Newsfeed or Instagram Feed are an increasingly common feature of social media platforms [3, 11, 13]. These systems often change more frequently than traditional technological systems [13], and that change has sometimes been met with public resistance from users. Prominent examples include early backlash when Facebook’s Newsfeed was introduced [15] (and expanded [1]) as well as more recent negative reactions to changes in Twitter’s Timeline [26] and Instagram’s feed [22]. These negative user reactions

to algorithmic change may indicate larger problems extending beyond these commercial platforms such as inaccurate user understandings of how the systems work, both before and after potential changes (e.g., [5, 10, 11]), as well as designer misunderstandings of what users see as important aspects of platforms (e.g. [25]).

User misunderstanding of algorithmically-driven curation systems has been documented [5, 10, 11, 29], with the implication that more accurate understandings might lead to increased user agency and success in achieving goals, as well as increased user trust in these systems [10, 21]. This could also help address emerging skill and literacy gaps around algorithms (e.g., [10, 13]). Moreover, increased sensitivity to how users (mis)understand systems may help designers avoid negative reactions to platform change.

As it stands, however, we know little about why users resist algorithmic changes or the extent to which users understand the systems they are resisting. Existing theories that aim to understand acceptance and rejection of new technologies and changes focus primarily on relatively slow-changing organizational settings where user agency is limited by a lack of outside choice (e.g., [2, 8, 24]). These theories could be valuable in understanding reactions to algorithmic change, but are difficult to apply directly to the unique situation of constantly-updated, algorithmically-driven social media feeds, where users are not employees, have numerous other platform options, and often have a central and strong voice in establishing popular sentiment with respect to the system.

Recent research suggests that folk theories, which capture a user’s working understanding of system operation [12, 28] and can act as a high-level frame for shaping user expectations [7, 27], may be a useful window into understanding user resistance to algorithmic change [5, 10, 12]. User expectations are a key element of this problem, as reactions are likely driven by the degree to which a change fulfills or violates their expectations of the system [2, 24]. Moreover, user reactions themselves are a potentially valuable source of data in that they can reveal both latent folk theories and system expectations.

If we are to make progress in this area, it is important to understand the nature of user reactions to changes in algorithmically-driven systems (and especially their complaints), what folk theories users articulate in the face of algorithmic change, and how these folk theories are expressed through and potentially frame user reactions.

In this paper, we address these questions by examining an “in the wild” corpus of tweets in response to a 2016 BuzzFeed News article (falsely) claiming that Twitter would change the organization of its timeline from simple reverse chronological order to one based on algorithmically-determined relevance to each user [18]. This article prompted an outpouring of sentiment on Twitter using the #RIPTwitter hashtag [26], even though the described changes were not made to the platform’s feed algorithm.

We examine the entire English-language #RIPTwitter corpus via human and machine content classification. Our results suggest that: 1) user resistance to algorithmic change is primarily based on the potential violation of specific prior expectations set up by the user; 2) implicitly-expressed user folk theories regarding these offending algorithms are mostly nonspecific, in that they consider algorithms as external agents of change, but do in some cases rise to the level of theories about how the algorithms actually work; and 3) user folk theories regarding algorithms seem to be related to user reactions, such that more specific folk theories tend to be linked to more specifically-expressed reactions.

BACKGROUND

The #RIPTwitter tweets feature direct discussion of and reaction to algorithms, in which users likely draw on their folk theories of algorithmic operation. In this section of the paper, we review theories of technology acceptance and rejection with a focus on understanding the nature of resistance to new technology and on how users develop and articulate folk theories of how these systems operate.

Resistance to Algorithmic Change

The #RIPTwitter tweets can be seen as an immediate outpouring of possible reasons for accepting or rejecting changes to Twitter’s core content delivery mechanism. One useful theoretical model for considering these reasons to reject or accept is the expectation-confirmation model (ECM)¹. This model holds that continued use of information technology is determined by continually-confirmed user satisfaction, a product of the interplay between user expectations and the technology’s performance relative to those expectations [2]. ECM has been extended to take perceived ease of use and user enjoyment into account [31], as well as task/technology fit [25]. In all versions, expectations are key: continued use requires that users’ initial and, more importantly, continually-reevaluated expectations of a technology are met. In considering rumored Twitter changes, continued use of Twitter would depend on whether it continued to fulfill users’ expectations of it.

Expectation violation is also at the heart of resistance to technology change. As a baseline, a technology’s status quo serves as the basis for an expectation, which can potentially be violated by change. The violation’s negative impact can

be exacerbated by high switching costs and a user-perceived lack of agency in the decision [20].

Expectation violations directly threaten the user’s understood status quo, which Lapointe and Rivard [24] have identified as a key driver of resistance to new technologies. In their model of resistance to technology implementation, they posit a process in which users assess the match between the features of a new technology and the place that technology would occupy in their lives. If that assessment of the consequences of the new technology threatens the status quo, resistance begins.

As an example of how the ECM and related theories play out, we can briefly look at an example of user backlash against Facebook’s News Feed. The backlash centered around changes to both visibility and curation of content alongside the introduction of the News Feed feature, which regularly displayed content from other users, as a “home” page instead of the user’s own profile [15]. Users had prior experience and an established expectation that content would only be displayed to certain people on certain pages. The rumored change to a feed where any content was potentially displayed to all of a user’s contacts represented an abrupt overturning of the status quo for both information flow and privacy, and therefore was likely a significant expectation violation for many users. As such, the rumored change was resisted, to the extent that Facebook had to temporarily backtrack, publically apologize, reengineer the feed, and re-launch after taking user complaints into account.

In this prior case, and potentially in the case of #RIPTwitter, the new, algorithmically-driven system appears to threaten a status quo. However, the Facebook example took place before the word “algorithm” had entered common public lexicon, and #RIPTwitter allows us to look at a response predicated directly on responding not just to an algorithm’s effects, but to the introduction of the algorithm itself. In turn, this allows us to examine if the documented focus on expectation violations in the Facebook case was also present for #RIPTwitter’s instance of resistance to algorithmic change. Accordingly, we asked:

***RQ1:** What reasons for resistance to the idea of algorithmic change did participants in #RIPTwitter express?*

Understanding the Changing Algorithm

As noted above, user expectations of technology may be shaped by perceived use cases [25], but they are also potentially shaped at a higher level by user understanding of the system’s inner workings [27].

Prior work has shown that users of algorithmically-driven platforms may not be aware of their algorithmic nature. Those that do have some idea of what is going on often understand through folk theories which are, in general,

the framework of the core Twitter technology; as such, continuance is a more appropriate lens with which to examine #RIPTwitter.

¹ We also considered longstanding theories such as the technology acceptance model [8]. However, though algorithmic curation could be considered a new technology, it was introduced to users within

oversimplifications and often inaccurate [10, 11, 29]. However, this lack of technical knowledge is not necessarily a barrier to acceptance or resistance.

Several related lines of work have established that user folk theories can function in the place of specific knowledge of what an algorithm does [12], allowing the individual to respond to an algorithm's perceived behavior, an arrangement that Bucher has labeled the "algorithmic imaginary" [5]. In this manner, folk theories, acting in place of actual technical knowledge, can affect how users respond to a system. Similarly, Orlikowski and Gash have established that users respond to change in technology based off of their own assumptions, expectations, and knowledge of how a system works [7, 27]. In this case, user folk theories constitute those assumptions, expectations, and knowledge, and essentially act as the contextual frames through which users interact with a system. As such, examining these folk theories can give us insight into how users' working understandings of algorithmic systems might affect the expression of user resistance.

In considering what constitutes a "folk theory," we started with Gelman & Legare's [12] concept of "intuitive causal explanatory theories that people construct to explain, interpret, and intervene in the world around them." We found this definition too restrictive, as it assumes strict causal reasoning. According to Orlikowski & Gash, the technological frames that individuals use to understand and react to technology are based not just on knowledge, but assumptions and expectations as well, especially those concerning consequences, and constitute an exercise in sensemaking [27]. In acting as a type of sensemaking frame, we argue that a folk theory can call on relevant, yet informal, abstract, or partial ideas, as noted by Keil [19], and need not be mechanistic to express an assumption about the consequences of a technology. As such, we adopt an expanded definition of "folk theories" as *intuitive, informal theories that individuals develop to explain the outcomes, effects, or consequences of technological systems, which guide reactions to and behavior towards said systems*. This includes causal models of how an algorithm might work, as well as opinions and attitudes about possible consequences of how it operates.

Prior work by Eslami et al. on directly eliciting folk theories regarding the Facebook Newsfeed found that users had a variety of highly specific folk theories, ranging from curation on the basis of personal engagement with other accounts, content, or types of content, to attempts by the platform to balance out types of friends or content [10]. These can be considered specifically causal speculations on what algorithms do, with an additional theory that attributes curation to "randomness." Similarly, Rader and Gray found evidence of folk theories that pointed specifically to causal mechanisms, such as a platform's attempt to provide "relevant" or popular content [29]. In both cases, the folk

theories elicited were usually developed to the point of the user making specific, mechanism-level causal claims.

#RIPTwitter is a somewhat different situation. In contrast to the above studies, expressed folk theories are based on speculation as to how a system will work rather than how a system does work. However, this type of implicit folk theory can still be a window into how users believe a system operates (or will operate) and how that belief guides behavior and understanding [12]. In fact, it may actually bring us closer to an understanding of how these systems are operationally understood "in the wild" [16]. In turn, this may allow us to determine the extent to which reactions to change in algorithmically-driven systems are framed by and subsequently expressed in user folk theories [5, 12, 27]. Accordingly, we asked:

RQ2A: *What folk theories of algorithmic influence over Twitter were displayed by #RIPTwitter participants?*

RQ2B: *Do certain types of reactions to algorithmic change reflect the expression of different folk theories?*

METHODS

To explore these questions, we used a content analysis of the entire English-language corpus of #RIPTwitter tweets. As this corpus represents a single event or case, an extensive, in-depth look at the whole phenomenon is essential, per [33]. As such, we used human content coding to classify a subset of the corpus for both quantitative [32] and thematic [34] analysis, and then used supervised machine classification to broaden our analysis to the whole corpus, per [30].

Data Collection

Our data set consisted of 102,827 unique English-language tweets that used the #RIPTwitter hashtag. These were collected via the Twitter REST API during the five-day period when the hashtag was active after the inciting article ([18]) was posted, February 5 - 10, 2016. This collection resulted in an initial corpus of 250,482 tweets. The initial corpus was then filtered in R, using an API-provided flag to remove retweets and the *textcat* package to identify and remove non-English tweets. Though this data was exclusively pulled from public Twitter accounts, we have treated it as if it were private throughout the coding and analysis process, and only reproduce tweets in this article that are still posted and public as of submission.

As it was not possible for human coders to review all 102,827 tweets, we randomly selected a subset of 10,050 tweets to act as a first point of analysis as well as the ground truth training data for the subsequent supervised machine learning model.

Data Set Characteristics

As our data come entirely from one platform during one incident, we suggest caution in interpreting our results. Twitter's users include about 23% of all online adults, and demographically trends younger, more urban, educated, and non-white than other social media platforms [9]. We were focused on studying the reactions of individual Twitter users,

not the reactions of groups, and as such we examined our dataset to check that it was appropriate for this individual focus. We found no single individuals who dominated the #RIPTwitter discussion, suggesting that our data represent all #RIPTwitter participants, and not just a few loud individuals. In the hand-coded set, 9,063 individual accounts are represented, with none comprising more than 0.004% of the set; in the full corpus, there are 71,171 accounts, with none comprising more than 0.003%. We also checked to see if the corpus contained back-and-forth dialog between users or long threads from individuals, both of which would threaten the validity of our claims. Only 10% of tweets contain an @mention, and less than 0.01% of the tweets are from recurring users, suggesting very few conversations or threads. Thus, we believe the data is appropriate for our goal.

Coding Procedure

We iteratively and inductively developed a three-level (valence, reaction, algorithm theory) coding scheme through discussion between the lead author and three undergraduate research assistants acting as coders, per [32]. The entire scheme was also reviewed by the authors during the process.

The three coders applied the coding scheme, starting with a 4,572-tweet training phase in which tweets were coded in small blocks, compared and reconciled by all coders. Once agreement was consistently above 70% for all categories, an additional 5,478 tweets were coded by a single coder, with 20% of this second group subsequently cross-coded by an additional coder, which resulted in acceptable Cohen's kappa reliability scores (per [23]) for all levels of the scheme (valence: 0.85, reaction: 0.79, algorithm theory: 0.78).

Classification Procedure

To better capture the entire English-language #RIPTwitter phenomenon, we then applied machine learning to classify the remaining 92,777 tweets with the same coding scheme. This was executed in Python using NLTK and scikit-learn. We used the hand-coded set as ground truth to train a series of support vector machine (SVM) classifiers, a common classifier for text-based problems [30].

For each classifier, each tweet was processed through a pipeline that first normalized the text. Normalization included standard steps such as removing whitespace and punctuation; however, to preserve as much data as possible from the short tweets, and to account for frequent use of casual language, we preserved word lengthening (e.g., "nooooooo" as more emotive than "no") per [4, 17], and added a flag to indicate that a tweet was "shouting" if it was more than 50% uppercase letters. We also normalized URLs and direct @mentions. URLs were resolved to their endpoint to counter link shortening, and the top 30% of URLs were given unique flags, with the rest receiving a generic URL flag. Similarly, the top 5% of @mentions, plus a list of news organization handles, were given unique @name flags, while the rest received a generic @mention flag. Both of these cutoffs represent the point at which their respective distributions, both of which were long-tailed, flatten out.

After cleaning, the tweets were run through NLTK's Twitter-specific tokenizer, which preserves emoji and emoticons, and then a tf-idf vectorizer to extract unigrams and bigrams. Finally, we used a chi-square test to select the top 30% of features for use in the final classifiers.

Three SVM classifiers were trained on this ground truth set with kernel parameters optimized via grid search. The classifier for valence used a linear kernel and, to combat class imbalance, had synthetic minority-class data added to the training set via SMOTE [6]. The classifiers for reaction and algorithm theory used a radial basis function kernel. All classifiers were evaluated using 10-fold cross validation to produce mean F1 scores (valence: 0.83, reaction: 0.67, algorithm theory: 0.76).

RESULTS

In this section, we first establish the nature of the #RIPTwitter resistance by reviewing the reasons participants expressed for their resistance to Twitter's change. We then explore the nature of users' expressed understanding of algorithmically-driven systems by examining the folk theories gleaned from participants' expressions. Finally, we show a relationship between user reaction and user folk theories. Note that we present results from our hand-coded dataset side-by-side with the machine-classified set. Results for both sets are similar, though not identical.

Reacting to an Algorithmic Rumor

First, we asked what reasons #RIPTwitter participants expressed for resisting change to Twitter (RQ1). Before answering this question, however, we checked the overall emotional valence of the tweets to validate the implicit assumption that #RIPTwitter was an instance of resistance.

A Negative Reaction

To check the actual valence of #RIPTwitter, we classified the emotional valence of each tweet's overall expressed feeling towards the rumored changes. Tweets were classified as "positive" if they expressed positive emotional reactions, such as pleasure, excitement, or gratitude, e.g.:

```
3/ In short, I think any algo-timeline (as a  
new choice to surface great Tweets) is a great  
move. The best is yet to be. #RIPTwitter
```

Tweets were classified as "negative" if they expressed implicit or explicit negative emotional reactions towards the rumored change, such as frustration, sadness, or anger, e.g.:

```
#RIPTwitter I hate Facebook because of this.  
Why are you doing this to me. You got rid of  
the star and now the timeline.
```

Tweets were classified as "neutral" if the overall sentiment of the tweet committed to neither of these emotional reactions, and remained even-keeled, e.g.:

```
Change is usually good. But change can also  
suck. We'll see which one this ends up being.  
#RIPTwitter
```

Supporting the assumption of a predominantly negative reaction, we found that 68% of the tweets (6,826) in the

Reaction	Definition	Example	Percentage (N)
Generalized Anger / Frustration	The user is expressing a simple disapproval, admonishment, accusation, or other expression of anger or frustration, without specifics.	<i>Fuck your algorithms @twitter #Riptwitter</i>	31.2% (2,587) 27.1% (23,484)
Metacommentary	The user is discussing #RIPTwitter itself as a phenomenon.	<i>Ironically, we're live-tweeting the death of Twitter's live feed. Do it now, while you still can! #RIPTwitter</i>	14.5% (1,206) 16.3% (14,132)
Explicit Expectation Violation	The user has identified some key worth or utility regarding Twitter that is now potentially being removed, replaced, or otherwise violated.	<i>Just don't, @twitter. We like chronological tweets. And live tweeting. And TWITTER BEING TWITTER. Leave it alone. #RIPTwitter</i>	12.8% (1,062) 15.3% (13,207)
Platform Duplication	The user is primarily commenting on Twitter attempting to copy another platform.	<i>Hey @twitter, If we liked Facebook, we'd be on Facebook. @jack #RIPTwitter</i>	11.9% (996) 14.7% (12,698)
Resignation	The user accepts that the rumored changes will inevitably happen, and may be searching for a new platform of choice.	<i>Looks like Twitter is going to turn into garbage. Was only a matter of time before the last good social media was ruined... #riptwitter</i>	8.9% (745) 11% (9,552)
Confusion	The user expresses genuine confusion over what #RIPTwitter is about.	<i>#RIPTwitter Wait what's happening?! Can someone explain?!?</i>	6.2% (512) 5.9% (5,160)
Platform Attachment	The user is primarily expressing general attachment or loyalty to Twitter as the user understands it (in actual or idealized form).	<i>Twitter you're perfect just the way or are. You don't have to change for anyone. Stay beautiful #RIPTwitter, please stay beautiful #RIPTwitter</i>	4.7% (394) 3.7% (3,217)
Ignoring Feature Requests	The user has called out Twitter for even considering the rumored changes when other critical requests (e.g., spam, abuse controls, edit button) have not been met.	<i>people: we want an edit button twitter: a what? people: an editwitter: algorithmic timeline. got it! :) #RIPTwitter"</i>	4.2% (348) 3.9% (3,343)
Wait and See	The user is inconclusive and either holding off judgement or posing conditional terms for acceptance or rejection of the rumored changes.	<i>Not willing to say #RIPTwitter until after the roll-out. If I have an option to go back to chronological, then no harm done.</i>	3.4% (281) 1.6% (1,394)
Economic Motivation	The user has identified the rumored changes as explicitly economically motivated, a "cash grab."	<i>Relevance=What is "relevant" to Twitter making \$. It's ALL \$ making. #SadButTrue #RIPTwitter</i>	1.5% (126) 0.3% (294)
Fixing the Problems	The user has identified the rumored changes as a potential fix for existing problems.	<i>What's with #RIPTwitter? Timelines are full of garbage. Surfacing the interesting bits would be a welcomed change.</i>	0.6% (48) 0.1% (53)

Note: Top percentages indicate results for hand coding; bottom percentages indicate results for machine classification. Tweets classified as off-topic/spam or purely informational (15.9% of the full corpus) are excluded, as they express no actual reaction.

Table 1: Classification categories for user reactions to the rumored algorithmic changes to Twitter.

hand-coded set had a negative valence, while 31% (3,146) were neutral. Only 77 tweets (0.77%) in the hand-coded set were coded as positive. The results from the SVM-classified entire corpus follow a similar pattern: 74.2% negative (76,327), 24.4% neutral (26,254), and 0.2% positive (245).

To ensure #RIPTwitter was not a negatively-valenced fringe conversation in a broader response to this rumor, we also looked for more positive responses by examining tweets from this period with hashtags often used to praise Twitter: "#goodjobtwitter" and "#thankstwitter." However, we found little evidence: "#thankstwitter" was not used during this time, and "#goodjobtwitter" was used rarely and often

sarcastically. We are thus confident that #RIPTwitter and the overall reaction to rumored Twitter changes were negative.

Violating the Status Quo

After verifying the negative overall valence of #RIPTwitter, we sought to better understand what reasons users might have for resisting by examining how participants expressed their resistance (RQ1). Overall, user reactions paint a clear picture of resistance prompted by perceived violations of their expected status quo, expressed in a variety of ways.

Our inductive coding scheme resulted in 11 distinct types of reaction messages (see Table 1). As we examined the tweets in each category, we realized that the categories seemed to

differ from each other in the level of detail with which participants described their expectations, perceptions of the status quo and potential violations of those expectations.

Based on this observation, we created a rough ordering of categories by the level of detail that typified tweets in that category. This ranges from the explicit mention of both a status quo expectation and how it is violated (*explicit expectation violation*) to the implicit assumption of the entire platform as currently constituted as the status quo and any change as a violation (*generalized anger/frustration*). In between these extremes lie categories that vary in the extent to which detailed descriptions are needed to express that type of reaction. We draw on this rough ordering in our analyses, but note that – apart from the clear endpoints – we cannot make fine-grained distinctions between categories. Note that two reaction types (*metacommentary* and *confusion*) did not fit this description pattern, so were not part of this exercise.

The clearest illustrations of perceived threats to user expectations around the status quo were found in the *explicit expectation violation* tweets. As Table 1 shows, these tweets mention the rumored changes as endangering specific features or the core functionality or “twitter-ness” that makes Twitter, Twitter:

```
I love Twitter because it's Twitter. Reordering
the timeline would be well... NOT Twitter.
#RIPTwitter
```

Explicit expectation violations were coded only when an expectation could be gleaned from the tweet, suggesting that the user had a specific use case for the technology, and/or a distinct place for that technology in their lives. In some cases, these expectations could be about what makes Twitter unique in a larger sense. For example, users may have come to expect that Twitter would provide them with certain types of social utility and functionality, e.g.:

```
I like twitter for a reason. It's a great
platform to share views and opinions. Please
don't take that reason away from me.
#RIPTwitter
```

In other cases, this was a very specific use case, where Twitter has an inherent value in fulfilling or completing some key task for which the user relies on the platform. Live-tweeting, in particular, was frequently mentioned:

```
Just don't, @twitter. We like chronological
tweets. And live tweeting. And TWITTER BEING
TWITTER. Leave it alone. #RIPTwitter
```

Similarly, Twitter’s functionality as a key source for breaking information was called out as a valued part of the current status quo, for both civic and personal reasons:

```
Twitter - "We'll get to the earthquake in a
second. But first, here are some cat pictures
we think you might like." #RIPTwitter
```

```
I use Twitter for live updates on things -
sports. That's why Twitter is amazing. What's
the point of a fucking algorithm????
#RIPTwitter
```

In all these cases, participants articulated an already-formed expectation of how Twitter “should” perform relative to their use case. There is a sense in which this can be seen essentially as an ad-hoc assessment of task/technology fit. This direct expression of an expectation that might be violated shows a clear sense of a (perceived) status quo among participants. To them, Twitter *is* or *is for* a distinct part of their online lives.

This understanding of an imperiled status quo (and, therefore, imperiled platform value) can also be found in many of the less frequently observed categories, such as platform duplication, resignation, platform attachment, economic motivation, and fixing the problems, totaling slightly over a third of the entire corpus (33.6% hand coded / 37.9% machine classified). Definitions and examples of these categories are in Table 1, but ultimately, despite ostensibly centering on other complaints about the system (e.g., an unseemly level of monetization in the case of *economic motivation*), they all also express that there is an implicitly or explicitly recognized status quo on Twitter that users see as threatened by the coming change. The more explicit cases, such as *platform duplication*, specifically call out Twitter as a distinct platform, and express an expectation that it will remain unique, e.g.:

```
Why can't Twitter just be fucking unique
instead of copying Facebook #RIPTwitter
```

The more implicit coding categories, such as *platform attachment*, recognize that there is something special about Twitter, sometimes in a very personal way, without necessarily being specific, e.g.:

```
I'm sorry it took something like this for me to
say it, but @twitter has been one of the only
joys in my life. #RIPTwitter
```

In these implicit cases, we still see evidence of expectations around Twitter’s nature that are seen to be threatened. These may not be expectations in the sense discussed in the literature reviewed above, but it is clear that these concerns are important and influencing the user’s reactions in similar ways. We will return to this point in the discussion.

Even tweets classified as *generalized anger/frustration*, which are by their nature less focused on specific platform features than many of the other categories, have an implicit concern for a status quo. As noted in the first line of Table 1, they are often about venting and admonishment, or dark, disparaging humor. They also turn to outright insult to Twitter as a company and its CEO in particular, sometimes rising to the level of threats. Individuals such as this one seem to seek an outlet for their frustration:

```
Just found out about the Twitter algorithm
thing. Here's my letter to Twitter, in which I
voice my full opinion:
FUCK YOU
#RIPTwitter
```

These tweets may not express specific reasons for being upset, but it seems clear that these individuals believe the

Code	User Folk Theory	Example	Percentage (N)
<u>Abstract Theories</u>			
Generic	Algorithm will affect the timeline in some nonspecific way; algorithm is largely defined as an external force.	<i>Algorithms ruin everything #RIPTwitter</i>	33.5% (778) 39.5% (8,769)
Opposition	Algorithm will oppose the current status quo of a chronological timeline; algorithm is largely defined by what it is not.	<i>it's called TIMELine not ALGORITHMICline #RIPTwitter</i>	20.7% (713) 32.6% (7, 237)
Comparison	Algorithm will operate exactly as another platform does; algorithm is defined exclusively as another platform's whole way of operating.	<i>These proposed twitter changes sound awful. I don't need another facebook-like feed. #RIPTwitter</i>	4.9% (113) 1.8% (401)
<u>Operational Theories</u>			
Popularity	Algorithm will display content based on overall popularity of content or specific accounts (e.g., celebrities), either across the platform or a smaller network (e.g., geographic area).	<i>Showing tweets in order by most popular, LOL some tweets will never be seen. #RIPTwitter</i>	11.6% (270) 10.3% (2,277)
Platform Directed	Algorithm will display what Twitter as an entity wants users to see.	<i>Only huge accounts will be in timelines. And advertisers. Just like Facebook. You'll see only what Twitter says you'll see. #RIPTwitter</i>	11.5% (266) 8.8% (1,954)
Relevance	Algorithm will use some user-based metric (e.g., clicks, page views, strength of ties, etc.) to select content that it believes the user wants to see.	<i>Let me decide what I want to see not what you think I want to see. #RIPTwitter</i>	7.8% (181) 6.9% (1,539)

Note: For percentage, top numbers indicate results for hand coding; bottom numbers indicate results for machine classification.

Table 2: Classification categories for user folk theories about the rumored Twitter timeline algorithm

likely impact of the rumored change would be negative. In many cases, these participants are essentially objecting to change of any type. This can be read as a very general violation of an ill-defined expectation that things will remain as they are now. This is another illustration of how people's attachment to the nature or features of a system cause concern that changes to those features might threaten the value they derive from the system. Where the source of this value is unclear, people seem to see any change as a possible threat to that value.

Theorizing the Rumored Algorithm

To better understand what might be framing or influencing user resistance, we also investigated the folk theories expressed by #RIPTwitter participants about the rumored algorithm (RQ2A). In particular, we wondered if specific types of folk theories would influence the way people discussed the rumored changes, and if this could help us understand people's perceptions of how the system works and why/how they derive value from it.

We found six distinct types of user folk theories (see Table 2). Overall, it appears that the folk theories expressed via #RIPTwitter do not show detailed, causal theories of how algorithmic curation might work. However, we did find a diversity of more general, high-level theories which seem to indicate differing levels of user understanding. That difference allows us to group the theories into two broad categories: *operational theories* and *abstract theories*.

Operational theories demonstrate a specific understanding that there are some criteria by which an algorithm must make curation decisions, and comprised a minority of all expressed theories (30.9% hand coded / 26% machine classified). As shown in Table 2, operational theories included: content popularity, Twitter's internal priorities, and some formulation of what is deemed relevant for the user. For example, these users implicitly point to popularity/platform priorities and relevance as decision criteria, respectively:

Your TL will be filled with only popular tweets, ads and promoted accounts :(#RIPTwitter

Appears Twitter will be using an algorithm to place tweets IT feels you'd wish to see before others. Seems very wrong. #Uneasy #RIPTwitter

Though most operational theories lack detail on how an algorithm works, they do express a baseline understanding of algorithmic curation.

In direct contrast, abstract theories, which include the generic, opposition, and comparison classifications, account for most of the theories expressed. They do not include specific attempts to theorize how an algorithm might actually operate. Instead, they rely on a more general sense that an algorithm is something that will, in turn, cause something to happen to the Twitter timeline: just generically "something" in the case of *generic*, something different than the status quo in the case of *opposition*, and something similar to another platform in the case of *comparison*. For example, these users

User Reaction	Abstract Folk Theories			Operational Folk Theories		
	Generic	Opposition	Comparison	Relevance	Platform Directed	Popularity
Generalized Anger / Frustration	438.48*	307.62*	39.99*	80.16*	1.79	0.76
Metacommentary	219.31*	52.34*	14.92*	85.27*	9.97	15.06*
Explicit Expectation Violation	967.27*	384.35*	54.89*	450.82*	28.60*	41.60*
Platform Duplication	5.22	13.48*	1385.19*	27.88*	12.07*	34.11*
Confusion	62.60*	14.60*	2.68	13.48*	5.52	7.53
Resignation	1.04	0.01	21.15*	62.45*	3.75	69.00*
Platform Attachment	6.22	79.93*	0.35	14.26*	7.39	26.72*
Ignoring Feature Requests	123.88*	14.03*	2.47	8.64	40.04*	37.01*
Wait and See	11.43	4.55	1.78	14.98*	11.38	12.39*
Economic Motivation	0.64	15.01*	0.37	5.21	151.47*	0.66
Fix the Problems	0.07	0.00	0.38	0.14	0.40	0.30

Note: * = statistically significant after Bonferroni correction, $\alpha = 0.00075$; red/dark cells indicate combinations occurring less than expected by chance, green/light cells indicate combinations occurring more than expected by chance; white cells not significant.

Table 3: Chi-Square tests of significant differences between expected and observed cell counts across user reactions to algorithmic change and user folk theories; based on machine classification of the #RIPTwitter English-language corpus.

are aware of the algorithm primarily as a generic “other” and as a force that will disturb the status quo, respectively:

Twitters bout to be ran by an algorithm
#RIPTwitter

Twitter don't change your algorithm, we like it
just like this..... #riptwitter

Ultimately, only a minority of the total corpus expressed any level of algorithmic theory, as is to be expected based on previous work. In the hand coded set, 23.1% of tweets (2,321) expressed a folk theory; in the machine classified set, 21.61% of tweets (22,239) expressed a theory. Overall, these results suggest that where responses expressed a theory, they were diverse and reflected diverse levels of understanding.

Matching Theory to Reaction

Finally, we investigated whether certain types of reactions to algorithmic change expressed different types of user folk theories (RQ2B) in order to better understand how these theories might frame or influence reactions to change.

At a high level, there does appear to be a relationship between specific type of algorithm theory expressed and reaction to algorithmic change, as chi-square tests indicate the two variables are not independent (hand coding $\chi^2 = 788.39$, $df = 50$, $p < .0001$, $N = 2,297$; machine classified $\chi^2 = 4901.07$, $df = 50$, $p < .0001$, $N = 21,506$).

To answer the question in more detail, we used a contingency analysis via post-hoc single degree-of-freedom chi-square tests with a Bonferroni correction applied (corrected $\alpha = 0.00075$). This analysis suggests that the different levels we found in the first two sections of results are linked, such that more specifically focused types of reactions are more likely to express more specific folk theories, and less specifically focused types of reactions are more likely to express less specific folk theories. This suggests in turn that more detailed

levels of algorithmic knowledge, as expressed through folk theories, may allow or prompt more detailed, and therefore actionable, expressions of resistance. This may seem intuitive or obvious at first, but this isn't the case. It would theoretically be quite reasonable to react with a specific feature in mind, but describe it with an abstract theory (such as the hypothetical “The timeline makes Twitter what it is. Algorithms will destroy it”) and vice versa.

The more specific, or operational, folk theories (popularity, platform directed, and relevance) we found in RQ2A, are expressed through more specific reactions to algorithmic change more frequently than would be expected by chance. As noted in Table 3, explicit expectation violations, arguably the most specific user reaction type as it requires a specific use or property of the platform to be violated, more frequently express all three types of operational theory than would be expected by chance. Additionally, economic motivation reactions appear to frequently express a platform directed algorithm theory (68.52% hand coded; 45.1% machine classified), as users are specifically theorizing a causal mechanism that primarily benefits the platform in economic terms. Similarly, resigned reactions appear to express a popularity algorithm theory significantly more frequently than would be expected by chance (30.17% hand coded; 17.62% machine classified), which coincides with a common exasperated attitude seen in these tweets, e.g.:

Bro, I'm done. It's all just gonna be a big popularity contest. #RIPTwitter

In direct contrast to this, the abstract algorithm theories (generic, opposition, and comparison), were expressed more frequently through less specific reactions than would be expected by chance. For example, platform duplication reactions appear to frequently express a comparison algorithm theory (35.8% hand coded; 13% machine

classified), as users are reacting in relation to another platform. Similarly, reactions classified as ignoring feature requests also frequently express a generic algorithm theory (62.79% hand coded; 63.27% machine classified), which coincides with a frequent view of the rumored algorithm as an ill-defined solution to a problem that Twitter doesn't actually have, e.g.:

```
MASSES: "We want more than 140 & edits"  
TWITTER: "Ok great, well just fuck up  
everything including your timeline, instead  
#RIPTwitter"
```

Importantly, the plurality of users that reacted with generalized anger/frustration appear to be expressing a generic algorithm theory (46.7% hand coded; 56.42% machine classified), significantly more than expected by chance. This suggests that the most abstract theories are often expressed through the least explicit responses to change.

In an interesting contrast to the generalized anger/frustration case, the plurality of users who reacted with explicit expectation violations appeared to be expressing an opposition-based algorithm theory, where the algorithm opposes the status quo (47.54% hand coded; 45% machine classified). As noted above, while opposition is not one of the operational algorithm theories, it does require a solid knowledge of the status quo. This suggests that, even in the absence of operational theories, more specific folk theories are expressed through more specific reactions to change.

DISCUSSION

By examining the English-language #RIPTwitter corpus, we have gained insight into the ways users express both their reasons for reacting negatively to algorithmic change and the folk theories with which they conceptualize the system itself. We found that user resistance is centered on violations of the user's expectations of a platform's status quo, and that this core concern is expressed in a variety of ways with varying levels of detail. We also found evidence of both operational and abstract user folk theories that vary in their level of detail or expressed user knowledge. Finally, we have shown a relationship between type of user reaction to change and level of detail in the accompanying algorithm theory, such that more detailed folk theories occur with more detailed negative reactions. All of this has implications for understanding what users value in social media platforms, how users conceptualize algorithmically-driven content curation systems, and how understanding one may be a potential window into understanding the other.

Expanded Expectations: A Personal Threat

Our findings regarding users' reactions to change present an opportunity to reevaluate and extend long-standing theories for the new realities of social media platforms. As noted in our results, the primary component of negative participant reactions was the possibility of a changed status quo that might violate expectations. At its core, this is consistent with both the ECM [2, 31] and Lapointe and Rivard's resistance framework [24], as they center around continued

confirmation of expectation satisfaction to keep users using a system. There is continued utility to understanding rejection (and acceptance) of systems using these theories, but to do so we must expand them to include new elements, including the additional expectations being violated.

Overall, we found expectations were focused on each user's value for the system, essentially an ad-hoc form of task/technology fit [25], but our results show that value is derived from other sources as well. Where value from systems described by these theories in the past stemmed from functionality such as the ability to complete work tasks or communicate effectively within organizations, #RIPTwitter participants showed that they derived value from work, personal, and community tasks, as well as much more abstract community dynamics. As we noted in our results, these community dynamics can include a simple sense of belonging, or of Twitter being a safe space for self-expression or emotional support. Task/technology fit here is less directly about supporting a task, but rather about whether changes to information delivery mechanisms within a community would impact how that community functions.

Compared with sources of system value in these theories, the more abstract sources of value, or system expectations that we saw do not require a discrete violation episode to seem potentially threatening to users. As we saw in our generalized anger/frustration results, even the very prospect of any change at all provoked intense, vulgar, and sometimes personal backlash for users that were valuing a perceived characterization of the system as a whole instead of an individual component. We saw users value, and therefore set expectations of, the system whether or not they understood their actual reasons for using a system. They were willing to draw their proverbial lines in the sand over what may seem like small changes to designers or researchers.

As such, future work on expectation violations and user resistance can still derive value from models like the ECM as a starting point, but should be cognizant of the expanded set of circumstances that might trigger "expectation violations" for users.

User Folk Theories: Few Specifics, Yet Deep Diversity

Our findings regarding users' folk algorithm theories present an interesting look into not just how users are conceptualizing the rumored timeline curation algorithm, but how they might be defining "algorithm" to begin with. As noted in our results, we found both abstract and operational folk theories reflecting two notions of algorithms: as an other or interloper, and the algorithm as a process requiring decision criteria, respectively.

This is a different gradation of algorithmic awareness than has been previously considered in empirical work, as prior studies (e.g., [5, 10, 11, 29]) have focused on operational theories, defining algorithmic awareness as starting at what we call operational folk theories. Our findings indicate that a finer-grained definition of algorithmic awareness, which

includes the abstract folk theories we have found, may be necessary to capture the full scope of what people understand about algorithms. This allows algorithmic understanding, as a concept, as well as future studies that deal with algorithmic understanding, to consider the perspectives of users who are newly or vaguely aware of algorithms as a concept, but have not progressed to theorizing how the algorithms operate. This expanded concept of algorithmic awareness would also lend empirical support to theoretical work (e.g., [14]) which argues that “algorithm” is defined in a multi-faceted way that goes beyond the technical meaning to account for social science and colloquial use of the term.

One type of folk theory we did not find were the complex, mechanism-level theories found by Eslami et. al. [10] as well as in prior folk theory work (e.g. [5, 11, 29]); even the theories we labeled “operational” were less detailed than many of the theories they elicited. This points to an important methodological implication, as our studies use two different methods of folk theory identification with diverging results. These prior studies used direct elicitation methods from lab-based and interview scenarios, while our study used indirect, inferential analysis of text “in the wild.” This suggests that future research should take into account (and, potentially, directly compare) the potential differences between directly elicited and indirectly identified theories noted by [16] as the two methods may be revealing different sides of a user’s folk theory: a causal theory stemming from active cognition of the algorithm at time of elicitation for the former, and a more implicit, always-extant “theory in use” for the latter.

Contextual Frames: An Open Question

Finally, our findings regarding how user folk theories frame or influence reactions to change point to an important open question. We saw a relationship between these two concepts, such that more explicit folk theories are expressed through more explicit reactions. This suggests that the context of a person’s reaction to and sense of personal expectation regarding a system should be considered when investigating folk theories, and vice versa, but further points to the need for discussing the relationship between a folk theory, reaction to change, and a system’s value for the user.

We have reason to believe, as we have explained in the literature review and noted in our results, that folk theories act as a frame for user reactions. Orlikowski and Gash’s work [7, 27] indicates that, for technology in general, prior assumptions about how a system works (here, user folk theories) act as a contextual frame for how a user reacts to that technology (here, their reason for resisting algorithmic change), and our results, viewed through this lens, could be an indicator that this paradigm applies even outside of its original business IT adoption context. As it appears that folk theories set the frame in which algorithmic change is processed and either accepted or rejected, it may be essential to take the context of an individual’s algorithmic folk knowledge into account when studying reactions to algorithmically-driven systems.

However, we do not show a causal relationship, and it is also possible that the inverse is true. In that case, expressed folk theories could be highlighting a post-hoc attempt by users to rationalize their expectations of system value by forming an appropriate folk theory of how the system works on a technical level. For example, an individual who explicitly values and expresses the fact that their use case/value proposition for Twitter is being able to see exactly what is going on, right at this moment, could rationalize a folk theory of an upcoming algorithmic change which directly threatens that, such as a relevance theory which would completely upend the value proposition. The implication of this possibility would be a new opportunity to understand how folk theories are formed, and would somewhat mirror the expectations of direct elicitation methods (e.g. [12, 16]).

Limitations

As with any study, certain limitations merit caution when interpreting these findings. First, as #RIPTwitter was an event that individuals had to self-select into, we cannot be sure how the participants map to Twitter’s whole user base; however, we have, at the very least, provided a window into the folk theories and expectations of some of the platform’s most vocal power-users. Second, as noted in our methods section, this data is based on a single case on a single platform, and therefore is tied to the circumstances of that platform and case. Future work should look at similar incidents on other platforms, as well as seek out implicit user reactions and folk theories that are not in response to an isolated incident. Third, as this case is based around early resistance to a rumor of algorithmic change, future work should also examine user reactions and folk algorithm theories during later steps of the change process. These include during and after implementation, in situations where algorithmic change is noticed incrementally by users, and in situations with constant A/B testing. Finally, as the corpus used was exclusively English-language, future work should examine potential cross-cultural differences in these areas.

CONCLUSION

#RIPTwitter was prompted by a rumor from a BuzzFeed article. In the end, some small-scale features to promote relevant tweets were added to the Twitter Timeline, but the non-chronological algorithmic apocalypse prophesied by the community did not come to pass. However, examining this outpouring of sentiment has proven a useful exercise regardless, and provided a better understanding of both reasons why users resist algorithmic change and how their folk theory-based understandings of algorithmically-driven systems may relate to those reactions. This points a way forward towards understanding, assessing, and improving user knowledge of algorithmic systems.

ACKNOWLEDGEMENTS

We acknowledge partial support from the US National Science Foundation (IIS-1217143/003), valuable research assistance from Bennett Hensey, Minkyong Kim, and Justine Yucesan, and valuable insights from the anonymous reviewers and the associate chair.

26. Brett Molina. 2016. CEO Dorsey responds to Twitter timeline outrage. Retrieved from <http://www.usatoday.com/story/tech/news/2016/02/06/twitter-may-change-timelines-and-no-one-happy/79925572/>
27. Wanda J Orlikowski and Debra C Gash. 1994. Technological frames: making sense of information technology in organizations. *ACM Transactions on Information Systems (TOIS)*, 12, 2: 174-207.
28. Stephen J Payne. 2003. Users' mental models: The very ideas. In *HCI Models, Theories, and Frameworks: Toward a Multidisciplinary Science*, John M Carroll (ed.). Morgan Kaufmann, Burlington, MA, 135-156.
29. Emilee Rader and Rebecca Gray. 2015. Understanding User Beliefs About Algorithmic Curation in the Facebook News Feed. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 173-182.
30. H Andrew Schwartz and Lyle H Ungar. 2015. Data-Driven Content Analysis of Social Media A Systematic Overview of Automated Methods. *The ANNALS of the American Academy of Political and Social Science*, 659, 1: 78-94.
31. James YL Thong, Se-Joon Hong, and Kar Yan Tam. 2006. The effects of post-adoption beliefs on the expectation-confirmation model for information technology continuance. *International Journal of Human-Computer Studies*, 64, 9: 799-810.
32. Robert Philip Weber. 1990. *Basic Content Analysis*. Sage, Thousand Oaks, CA.
33. Robert K Yin. 2003. *Case Study Research: Design and Methods*. Sage Publications, Thousand Oaks, CA.
34. Yan Zhang and Barbara M. Wildemuth. 2009. Qualitative analysis of content. In *Applications of Social Research Methods to Questions in Information and Library Science*, Barbara M. Wildemuth (ed.). Libraries Unlimited., Westport, CT, 308-319.