User Attitudes towards Algorithmic Opacity and Transparency in Online Reviewing Platforms

Motahhare Eslami
University of Illinois at
Urbana-Champaign
eslamim2@illinois.edu

Kristen Vaccaro
University of Illinois at
Urbana-Champaign
kvaccaro@illinois.edu

Min Kyung Lee
Carnegie Mellon University
mklee@cs.cmu.edu

Amit Elazari Bar On
University of California, Berkeley
amit.elazari@berkeley.edu

Eric Gilbert
University of Michigan
eegg@umich.edu

Karrie Karahalios
University of Illinois at
Urbana-Champaign
kkarahal@illinois.edu

ABSTRACT

Algorithms exert great power in curating online information, yet are often opaque in their operation, and even existence. Since opaque algorithms sometimes make biased or deceptive decisions, many have called for increased transparency. However, little is known about how users perceive and interact with potentially biased and deceptive opaque algorithms. What factors are associated with these perceptions, and how does adding transparency into algorithmic systems change user attitudes? To address these questions, we conducted two studies: 1) an analysis of 242 users' online discussions about the Yelp review filtering algorithm and 2) an interview study with 15 Yelp users disclosing the algorithm's existence via a tool. We found that users question or defend this algorithm and its opacity depending on their engagement with and personal gain from the algorithm. We also found adding transparency into the algorithm changed users' attitudes towards the algorithm: users reported their intention to either write for the algorithm in future reviews or leave the platform.

KEYWORDS

Algorithmic Opacity, Algorithm's Existence, Algorithm's Operation, Transparency, Reviewing Platforms

ACM Reference Format:
Motahhare Eslami, Kristen Vaccaro, Min Kyung Lee, Amit Elazari Bar On, Eric Gilbert, and Karrie Karahalios. 2019. User Attitudes towards Algorithmic Opacity and Transparency in Online Reviewing Platforms. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland Uk.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3290605.3300724

1 INTRODUCTION

Algorithms play a vital role in curating online information: they tell us what to read, what to watch, what to buy, and even whom to date. Algorithms, however, are usually housed in black-boxes that limit users' understanding of how an algorithmic decision is made. While this opacity partly stems from protecting intellectual property and preventing malicious users from gaming the system, it is also a choice designed to provide users with seamless, effortless system interactions [5, 6, 12]. Still, this opacity has been questioned due to the biased, discriminatory, and deceptive decisions that algorithms sometimes make [9, 19, 37, 38].

One algorithm which has caused great controversy and dissatisfaction among users due to its opacity is the Yelp review filtering algorithm. Nearly 700 Federal Trade Commission reports have been filed, accusing Yelp of manipulating its review filtering algorithm to force businesses to pay for advertising in exchange for better ratings [26]. In addition to being opaque in operation, the Yelp review filtering algorithm is opaque in its very *existence*. That is, the Yelp platform interface not only hides how the algorithm decides what to filter, but also the fact that the review filtering algorithm is at work at all (Figure 1). When users discover this opacity, it can lead them to suspect the algorithm is biased, since it can appear the platform decided to intentionally hide the algorithm's existence or operation from them [22].

Recognizing the sometimes biased or deceptive effects of opaque algorithmic decision-making, policy-makers and academics alike have suggested robust regulatory mechanisms to increase the transparency, fairness, accountability, and interpretability of algorithms [9, 19, 38]. Informing these

CHI 2019, May 4–9, 2019, Glasgow, Scotland Uk

ACM ISBN 978-1-4503-5970-2/19/05...$15.00
https://doi.org/10.1145/3290605.3300724

regulatory and design proposals, researchers began investigating users' interaction with opaque algorithms in various online platforms such as social media feeds [8, 12, 14, 16, 36], service-sharing platforms [24, 28], and rating sites [15]. However, it is still not clear what factors are associated with users' different perceptions of such opaque algorithms.

Recent work has explored adding transparency into opaque algorithmic systems such as social feeds [35], grading systems [27], and behavioral advertising [13]. But what about algorithms whose opacity causes users to suspect bias or deception? Will adding transparency into such algorithms improve users' attitudes and increase user understanding, as hoped? Or might it instead occlude, as some have warned [2]?

In this paper, we address these gaps through two studies characterizing users' perceptions surrounding the Yelp review filtering algorithm (hereafter "the algorithm"), the factors associated with their perceptions, and the impact of adding transparency on users' attitudes and intentions. The first study collected and analyzed 458 online discussion posts by 242 Yelp users about the Yelp review filtering algorithm and its opacity in both *existence* and *operation*, identifying users' perceptions of this algorithm, and the factors associated with their perceptions and attitudes towards the algorithm. Building on this analysis, a follow-up study explored how adding transparency about the algorithm impacted users' attitudes. The study used a tool we developed, *ReVeal*, to disclose the algorithm's existence in 15 interviews with Yelp users to explore how users' attitudes changed.

We found that users took stances with respect to the algorithm; while many users challenge the algorithm and its opacity, others *defend* it. The stance the user takes depends on both their personal engagement with the system as well as their potential of personal gain from its presence. Finally, we found that when transparency is added into the algorithm, some users reported their intention to leave the system, as they found the system deceptive because of its opacity. Other users, however, report their intention to *write for the algorithm* in future reviews.

2 RELATED WORK

While algorithms exercise power over many aspects of life, they are often opaque in their operation and sometimes even in their existence. As Burrell discusses, this opacity stems from 1) corporate secrecy geared to prevent malicious users from gaming the system, 2) the limited technical literacy of regular users of these systems, and 3) the complexity of understanding an algorithm in action, even by its own developers [5]. In recent years, researchers have studied the opacity of algorithms and the impact of algorithmic opacity on users' interaction with algorithmic systems. For example, recent work studied users' of awareness of opaque social media feed curation algorithms [14] and users' folk theories on how algorithms work [8, 12, 16, 36]. Lee et al. investigated the benefits and drawbacks of powerful but opaque algorithms used in ridesharing services [28], while others analyzed the anxiety of users whose work practices were evaluated by opaque algorithms [24].

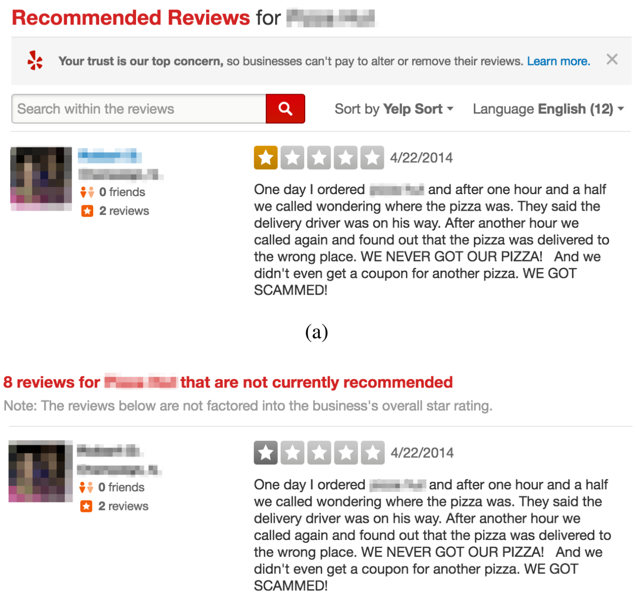
While algorithmic opacity can provide users with a seamless and effortless system experience, sometimes it can actually facilitate algorithmic decisions that are biased or deceptive. For example, Eslami et. al. found that the rating algorithm of a hotel rating platform (Booking.com) skews low review scores upwards (up to 37%) to present a better image of hotels. In this case, the nature of the algorithmic opacity, alongside misleading interface design choices, made it harder to detect the biased outcomes [15]. In another example, the opacity of the Uber rating algorithm led drivers to accuse Uber of deception – manipulating drivers' ratings in order to extract additional fees [41]. Accordingly, opaque algorithmic decision making has gathered considerable attention from legal scholars [17, 39].

Calls for Transparency

Such potentially biased or deceptive outcomes in opaque algorithmic systems have resulted in calls for transparency from users, researchers, activists, and even regulators to keep algorithms accountable [9, 21, 43]. These calls inspired researchers to study how adding transparency into opaque algorithmic systems impact user interactions with the system. Examples include investigating algorithmic transparency in algorithmically curated social feeds [35], news media [10], online behavioral advertising [4, 13] and recommender systems [23, 34, 40, 42].

Transparency, however, doesn't come without costs. While it might seem simply to educate users or help them understand decisions better, transparency can also introduce its own problems, particularly if not designed carefully [2, 7]. The wrong level of transparency can burden and confuse users, complicating their interaction with the system. [13, 27]. Too much transparency can disclose trade secrets or provide gaming opportunities for malicious users. For example, Yelp argues that the opacity of its review filtering algorithm is a design choice aimed to prevent malicious gaming [45]. Thus, while potentially helpful, adding transparency to opaque algorithmic systems requires finding the right *level* of transparency.

To find the right level of algorithmic transparency and how to design transparent interfaces, we first need to understand the types of algorithmic opacity, users' perceptions of and attitudes towards opacity, and the factors that shape these perceptions and attitudes. In this paper, we explore these questions through the case of the Yelp review filtering algorithm.



Recommended Reviews for [redacted]

Your trust is our top concern, so businesses can't pay to alter or remove their reviews. Learn more.

Search within the reviews

Sort by Yelp Sort ▾ Language English (12) ▾

[redacted]
0 friends
2 reviews

4/22/2014

One day I ordered [redacted] and after one hour and a half we called wondering where the pizza was. They said the delivery driver was on his way. After another hour we called again and found out that the pizza was delivered to the wrong place. WE NEVER GOT OUR PIZZA! And we didn't even get a coupon for another pizza. WE GOT SCAMMED!

(a)

8 reviews for [redacted] that are not currently recommended

Note: The reviews below are not factored into the business's overall star rating.

[redacted]
0 friends
2 reviews

4/22/2014

One day I ordered [redacted] and after one hour and a half we called wondering where the pizza was. They said the delivery driver was on his way. After another hour we called again and found out that the pizza was delivered to the wrong place. WE NEVER GOT OUR PIZZA! And we didn't even get a coupon for another pizza. WE GOT SCAMMED!

(b)

Figure 1: (a) A filtered review is presented as "recommended" to the user who wrote it while logged in (b) This review, however, presented for other users as a filtered review.

Case Study: The Yelp Review Filtering Algorithm

Users of online review platforms value their reviews greatly. For many users, reviews allow a creative voice, while providing them the most effective way to share satisfaction or disappointment with a service rendered, from life-saving medical treatment to a fast food meal. For business owners, reviews directly determine the business success and livelihood. Even a small, half-star change in a restaurant rating on Yelp can increase the likelihood of filling the seats by up to 49% [3].

However, while valuable, online reviews can be inauthentic and thereby potentially detrimental to both users and business owners. If reviews are written by business owners to promote their own business, they harm consumers; if written by competitors to undermine another business's reputation, they harm the business owner as well. To avoid this, Yelp employs a review filtering algorithm to decide which user-generated reviews on the platform are inauthentic or fake based on the "quality, reliability and the reviewer's activity on Yelp" [45]. Filtered reviews are not factored into a business's overall rating and are moved from the main page of a business to another page called "not recommended reviews".

While Yelp argues that its filtering algorithm is necessary, the opacity of this algorithm has caused controversies among users about the algorithm's bias and deception. Below, we describe two types of opacity in the Yelp filtering algorithm.

Opacity in Existence. Some algorithms are hidden on the interface, making it harder for users to know that they are the subject of an algorithmic decision-making process.For example, previous work has shown that many Facebook users were not aware of the algorithmic curation of the Facebook news feed [14]. While such opacity is often an unintentional consequence of design choices, it can be considered deceptive if the system appears to hide the algorithm's existence from users intentionally. In the case of Yelp, Yelp only reveals that a user's review is filtered when the user is logged out. When logged in, the user sees her filtered reviews under the recommended reviews of a business (as if unfiltered). So a user can only detect if reviews are filtered by looking for their own reviews for a business when logged out or logged in as another user. Figure 1 shows the difference: a review is presented as "recommended" to the user who wrote it while logged in (Fig 1-a), for other users this review is presented as filtered (Fig 1-b). This inconsistency in revealing algorithmic filtering of reviews can be deceptive to users.

Opacity in Operation. In addition to its opacity in existence, the Yelp algorithm is opaque in its operation. This opacity has led businesses to accuse Yelp of developing an algorithm that is biased against those that do not pay Yelp for advertising. In recent years, growing numbers of business owners have reported receiving calls from Yelp about its advertising – and that those who turned down the advertising noticed that long-standing positive reviews were filtered shortly after the call. Some even claimed that previously filtered negative reviews became unfiltered [22]. These complaints escalated into almost 700 lawsuits in recent years [1, 26], though all have been dismissed [20]. Yelp while denying the accusations of extortion [44], has argued that the opacity of its algorithm's operation is a design choice to avoid malicious gaming of the system [45]. However, it is unclear how users perceive and react to this opacity. To understand this, we asked:

RQ1: How do users perceive the **a)** *existence* and **b)** *operation* of the Yelp filtering algorithm and its opacity?

In addition to understanding users' perceptions of the opacity of the algorithm, we sought to understand why different users have different perceptions of the algorithm:

RQ2: What factors are associated with users' perceptions of the Yelp review filtering algorithm?

RQ1 and RQ2 aim to find users' *existing* perceptions of the algorithm and the factors associated with them; these questions, however, do not evaluate how users' attitudes towards the algorithm change after making some aspects of the algorithm transparent. This change is particularly important in opaque and potentially biased algorithmic systems where transparency has been suggested as a mechanism to establish more informed communication between users and the system. Therefore, we also sought to understand:

RQ3: How does adding transparency about the algorithm change user attitudes?

3 METHODS

We designed two studies to answer the proposed research questions: 1) a qualitative and quantitative analysis of 242 users' online discussions about the Yelp review filtering algorithm, and 2) an interview study with 15 Yelp users adding transparency about the algorithm via a tool that we developed.

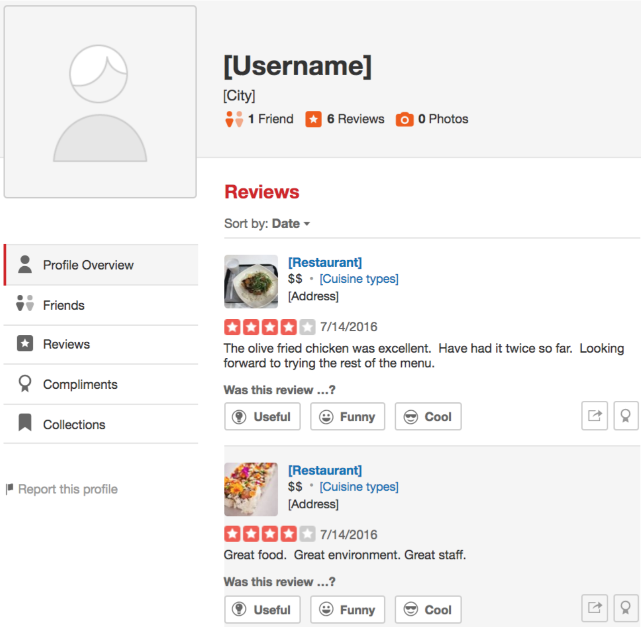
Study 1: Analyzing Online Discussions on Yelp

We conducted an initial investigation on Yelp, finding that Yelp provides users an "on platform" opportunity for discussion via forum pages. We searched for "review filtering algorithm" across Yelp (via Google search specifying a Yelp.com domain) to find posts concerning the algorithm, and how users discuss it. The search results included thousands of discussion posts, each up to nearly 10,000 words long. We selected the ten highest ranked forum pages discussing the algorithm's opacity in existence/operation and its potential bias and deception. In addition, since the Yelp algorithm changes over time, we expanded this set of discussions by adding the three top-ranked discussion pages in the search results for each year missing from the original set. The final set included 458 discussion posts by 242 Yelp users (the "discussants") from 2010-2017.

Data Analysis. To understand users' perceptions of the opacity in the algorithm's existence and operation (RQ1), we conducted qualitative coding on the discussion posts dataset to extract the main themes. A line-by-line open coding identified categories and subcategories of themes using Nvivo [32], and further revised these codes through an iterative process to agreement. We also conducted a quantitative analysis on the dataset to identify the factors associated with users' perceptions of the algorithm (RQ2). For clarity, details of both qualitative and quantitative analysis will be presented in the Results section.

Study 2: Adding Transparency Into the Algorithm

Analysis of the online discussion dataset provided a rich understanding of users' perceptions of the algorithm, yet most of the discussants were 1) aware of the algorithm's existence and 2) active enough on Yelp to participate in the forum. To analyze a more diverse set of users' perceptions, we conducted interviews with Yelp users to complement the results from the first study. That is, Study 1 and Study 2 complemented each other's results to answer RQ1 & RQ2, one with a population largely aware of the algorithm and one largely unaware. This avoided a bias towards previously aware users, given that prior work has shown different levels of awareness can result in different levels of engagement and behavior [14]. In addition, to understand how adding transparency into the algorithm influenced users' attitudes (RQ3), we developed a tool, *ReVeal* (Review Revealer), using which we disclosed the algorithm to users, showing them which of their reviews the algorithm filtered. To do so, we first collected a user's reviews from her public profile and inspected each review via JavaScript to determine if it had a "filtered" tag. The tool then highlighted the filtered reviews in the user's profile page using a gray background (Figure 2).



[Username]
[City]
1 Friend 6 Reviews 0 Photos

Reviews

[Restaurant]
$$ · [Cuisine types]
[Address]
7/14/2016
The olive fried chicken was excellent. Have had it twice so far. Looking forward to trying the rest of the menu.

[Restaurant]
$$ · [Cuisine types]
[Address]
7/14/2016
Great food. Great environment. Great staff.

Figure 2: The *ReVeal* tool shows users both their filtered and unfiltered reviews. Filtered reviews are highlighted with a gray background.

Recruitment and Participants. To find a set of typical users (not necessarily active or aware of the algorithm's existence like the discussants), we employed two methods of recruitment. First, we chose three random cities across the US and then five random businesses in each. For each business, we investigated both the recommended and filtered reviews. We contacted every user who had a filtered review for these businesses via the Yelp messaging feature. For each user who wrote a recommended review, we used our tool to check if they had any filtered reviews for other businesses. If so, we contacted them. Overall, we contacted 134 random Yelp users. Unfortunately Yelp perceived this as promotional/commercial contact and requested that we cease recruitment.

To reach more participants, we conducted local recruitment via flyers, social media posts, and other online advertising methods. In this approach, we restricted participants to those Yelp users who had written at least one review.

Via the above methods, we recruited 15 Yelp users (hereafter the "participants"). Nine had at least one filtered review (detected via our tool prior to the study). The participants had a diverse set of occupations including clerk, business owner, office administrator, librarian, teacher, student and retiree. The sample included nine women and ranged from 18 to 84 years old (40% between 35-44). Four participants were of Hispanic origin, 12 were Caucasian, three Asian and one American Indian. Participants had reached varying levels of education from less than a high school to a doctorate degree (about 50% with Bachelor's degree). Participants also had a varying incomes, from less than $10,000 to $150,000 per year. All received $15 for a half to one hour study.

The Interview. Participants first answered a demographic questionnaire including their usage of Yelp and other online reviewing platforms. We then assessed participants' awareness of the algorithm's existence by probing whether they knew a review might not be displayed on a business's main page. To do so, we first asked participants to log into their Yelp account. We selected their filtered review (or if they had multiple, chose a random filtered review) and asked them to show us where that review appeared on the business's main page. Since they were logged into their account, the review appeared at the top of the recommended reviews. Therefore, we particularly questioned them as to whether they thought other users would see the review in the same place. If they thought yes, we showed them where their review was actually displayed for other users, under the filtered reviews. Lastly, we asked if they had ever visited the list of filtered reviews for any business, and if they were aware of Yelp's practice of filtering reviews.

For participants with no filtered reviews of their own, we asked them to show us a random one of their reviews on the business' main page. We asked if they thought a user's review might not show up at all on that page, further probing for their awareness of the algorithm's existence.

After exploring users' existing knowledge of Yelp's review filtering, we asked them to share their thoughts about this practice. Next, participants compared the filtered reviews of a business (including their own reviews if they had a filtered review) with the recommended reviews, discussing their thoughts about why some reviews were filtered while the others were not. In doing so, we elicited users' folk theories about how the algorithm works. Users were also asked to discuss how they believed Yelp's interface should present the review filtering algorithm. Finally, we asked participants whether, in the future when they visit a restaurant, they would write a review on Yelp, and if so, whether they would change any of their review writing practices. The same qualitative method was used to analyze the interview results as was applied to the online discussions.

Stances **Algorithm**	Questioning (n)	Defending (n)
Opacity in Existence	24	0
Existence	33	32
Opacity in Operation	19	4
Operation	60	23

Table 1: The number of users who questioned or defended the algorithm's existence, operation, and its opacity in existence and operation.

4 RESULTS

The two studies found that users' perceptions and attitudes include taking strong stances with respect to the algorithm; while many users challenge the algorithm and its opacity, others *defend* it (RQ1). Table 1 provides an overview of the number of users who questioned or defended the algorithm's existence, operation, and its opacity. The stance the user takes depends on both their personal engagement level with the algorithm as well as the impact of the algorithm on their life (RQ2). We report the analysis of RQ1 and RQ2 by combining the discussions of both online users in Study 1 and interview participants in Study 2. Finally, we found that adding transparency into the algorithm changes user attitudes and intentions: some users reported their intent to leave the system, as they found the system deceptive because of its opacity. Other users, however, report their intent to *write for the algorithm* in future reviews (RQ3). We report results of both studies addressing these research questions, with participants from the online dataset labeled with *O#*, n_o and from the interviews with *P#*, n_p.

Qualitative Analysis: To analyze users' discussions in all the research questions, we conducted an iterative coding process. First, we read all the discussions several times and labeled them with preliminary codes. For example, for RQ1, these codes included a set of initial themes such as demanding for transparency, discouragement, freedom of speech, advertising bias, and demanding a change in the algorithm. We then analyzed the initial themes to find similarities and grouped them into final themes based on common properties. For example, in RQ1, we reached the main themes of "defending" and "questioning" the algorithm. In this section, we explain the themes for each research question in detail.

Perceptions of the Algorithm's Existence (RQ1a)

Questioning the Opacity in the Algorithm's Existence. Yelp's decision to hide the existence of its algorithm led many users (n=24: n_o=12 & n_p=12) to question Yelp's policy and design choices. First by critically stating Yelp's practices: "*Yelp gives it's users the illusion that one's reviews are all visible as they* *always remain from the users vantage point*"[1] (O1) but also by revealing to others how to uncover its existence: "*Yelp only shows me that my reviews are filtered if I am NOT logged in! If I am logged in, it shows my reviews for the restaurants as if nothing is filtered!!!*" (O92). Others focused on their emotions, sharing how the design choices increased their anger at finding their reviews filtered: "*so frustrating especially since Yelp doesn't indicate you're being filtered"* (O13).

Users questioned Yelp motives, suggesting that Yelp deliberately hides the review filtering process from users because "*they don't want their users to easily know when they're being suppressed [....] Kinda like they don't want their users to catch on that they've been poofed*" (O1).

A Benevolent or Deceptive Opacity? Prior work has shown that hiding the existence of algorithms can cause concern among users. The opacity of the Facebook News Feed filtering algorithm meant many users were unaware of its existence, causing initial surprise and concern [14, 36]. However, this opacity did not seem deceptive to Facebook users. In contrast, users labeled the Yelp review filtering algorithm "*sneaky*" (O2), "*deceiving*" (O125), even "*misleading and possibly censorship*" (O120) because "*it makes you seem like you are reaching out people more than you are*" (P10). The difference may again lie in design choices – in the fact Yelp users believed "*there's a little trickery involved there*" (P12) because "*if Yelp is so forthright, then why, oh why, is the 'filtered review' link buried at the bottom of the page behind a faint link that goes to a code to view?*" (O94).

Demanding Transparency in Existence: Feeling deceived by Yelp, users (n=14) demanded a "*full disclosure*" (O120) of the algorithm's presence through the interface design by putting the filtered reviews in "*PLAIN SIGHT*" (O120): "*I really would not have a problem with filtered reviews IF and only if they were located next to the reviews at the top and in a way that made it easier to know they are there. Being in gray at the bottom on the left where unless you knew Yelp well, you would never find nor click on them*" (O120). They also argued that Yelp has a responsibility to make the impact of filtering transparent to affected reviewers: Yelp "*should at least show [users] which of [their] reviews have been filtered*" (O5).

Users even compared Yelp to other platforms that provide more transparency in their filtering processes: "*I feel they should let you know if your review is being filtered. Because I know for other things like YouTube or things like that, if somethings filtered or restricted or whatever, they would tell you. So if you want to you could file a claim to remove it*" (P7). Highlighting the transparency provided and the opportunities for user control (e.g., a right to appeal), this user points to preferable alternative models for algorithmic filtering.

Questioning the Algorithm's Existence. Many users (n=33: n_o=25 & n_p=8) went beyond questioning the *opacity* in the algorithm's existence, to questioning the very existence of the algorithm. They argued that an algorithm should not control what users read: "*Yelp has a robot telling us how to feel about a business based on how the robot feels about our reviews... Good idea! Let's put this robot in charge of what we watch on TV and feed our children*" (O39). Instead, they asked Yelp to "*let the users be the judge*" (O77): "*I prefer my own intelligence instead of censors to help me to sort out 'absolute junk' from the reviews that matter*" (O73). In addition to framing the filtering algorithm as a tool for censorship, users fundamentally contested the need for the filtering at all; while they "*might be going to get tricked sometimes [by the fake reviews], [they] still don't want to get tricked in order not to get tricked*" (P6). They would "*rather just see everything and make [their] own decision than have a computer do it*" (P10).

Disengagement The presence of the algorithm made many users "*very discouraged about leaving reviews*" (O120) since they had spent time and effort to write reviews, but many of them ended up filtered: "*I am extremely saddened and disappointed to see that all my hard work has gone unnoticed due your software*" (O144). This became "*so frustrating*" (O13) for many users that they called writing a review on Yelp "*a simple waste of time*" (O164), expressing a sense of resignation and fatality: "*why did I even write a review then? It's just going to be filtered out*" (P15).

Users argued that this was particularly true for new reviewers, "*the filter robot is definitely a huge discouragement for new people*" (O95), and that one "*shouldn't have to write a TON of reviews to have a valid opinion*" (O92). In these cases, users closely tie *having* an opinion with being able to express and share it with others, thereby arguing that simply by filtering the filtering algorithm stifles them.

Freedom of Speech: In the same vein, some users argued that Yelp suppresses their speech: "*Yelp holding [itself] out as a public online forum and censoring people that are almost certainly legitimate reviewers*" (O9) "*totally defeats the whole premise of Yelp. The filter sucks the life outta creditability*" O(46): "*I don't want big brother Yelp [...] deciding what I can see. I suppose that's very American but [...] it's freedom of speech, freedom of information. I just don't like the idea that they would purport to be a public site that's crowd-sourcing reviews and would hide anything*"(P6). Users closely tie their ability to communicate with one another on these online platforms with their fundamental values.

Defending the Algorithm's Existence. While many users questioned the presence of a review filtering algorithm on Yelp, nearly as many users (n=32: n_o=31 & n_p=1) defended it. These users dismissed fears of censorship and deception, arguing that "*the filter is there to protect against malicious*

[1] All review text is presented exactly as it stands in the original, without notation of grammar, punctuation, or other errors.

content, for one, and also to ensure the content showing on Yelp is as useful and trustworthy as possible" (O18). These users even drew analogies to physical filters, arguing that Yelp with the presence of the filter "*is much safer. Otherwise, you are getting all that unnecessary tar and nicotine*" (O97).

Those who supported the filtering process elaborated on the numerous (and creative!) cases where content would need to be filtered, such as "*fake reviews from malicious competitors and disgruntled former employees*" (O89) and "*people who may not be 'real' (meaning, someone who came on here just to dump on an exboyfriend or pimp their friend's new business)*" (O90). They noted that not all reviewers are real people: "*there are a lot of ID's out there, 'bots' sometimes [...] with intent to provide reviews to beef up some businesses or downplay others*" (O71). Some pointed out that that even an otherwise upstanding reviewer might have a questionable review due to bribery: "*restaurants [...] have had Yelp flyers that tell you to leave a review and get a % off*" (O60). These users argue, therefore, that filtering is necessary. Often implicit in their examples are illustrations of the scale of this problem: Yelp needs to "*keep businesses from promoting themselves with dozens of fake reviews this way*" (O7), thereby requiring an *algorithmic* filtering intervention.

For such supporters, the algorithm was what distinguishes Yelp from other review websites: "*It's what sets us apart from other review sites and makes our content the best out there when it comes to local business reviews. Have you all read up on the filter in our blog posts?*" (O81). Thus, in discussions between those questioning the algorithm's existence and those defending it, the supporters argued that "*over time, [...] you will actually come to appreciate the filter for not drowning out YOUR thoughtful, carefully crafted reviews in a sea of irrelevant, profanity-laced and offensive tirades*" (O4).

Your reviews are not gone: Finally, in responding to the concerns of some users questioning the algorithm's existence, some users emphasized that no one's speech is permanently silenced by the algorithm, and that the algorithm constantly adapts and changes. For example, one participant noted that "*your reviews are not deleted and not permanently filtered, they could more than definitely show back up on the business page at any time*" (O64). Since the filtered reviews "*can still be found on the your personal profile page*" (O89) as well as "*among the 'not recommended' reviews*" (O71), the filtering algorithm is not really removing users' content. Furthermore, new reviewers can earn their place on the main page with time and effort: *"filtered reviews are often unfiltered after the reviewer accumulates more experience on Yelp*" (O48).

Perceptions of the Algorithm's Operation (RQ1b)

Of even more concern than the algorithm's existence was the algorithm's operation. Many users (n=140) engaged in discussions about the algorithm's operation, suggesting "folk theories" about how the algorithm works. Based on these folk theories, some users once again questioned the algorithm's operation (both in how it functions and the opacity of how it functions) while others defended it.

Questioning the Opacity in the Algorithm's Operation. When learning of the algorithm's existence, some (n=19: n_o=12 & n_p=7) questioned the opacity in the algorithm's operation, perceiving Yelp as being secretive about how its algorithm works: "*The unfortunate thing about Yelp is that if they decide to filter or delete a review or pic they are not specific as to why they did it*" (O15). This left users feeling powerless: "*thanks Yelp, you sure are mysterious and so are your Al Gore Rhythms. I certainly feel properly hazed & initiated. You can put away the Goat masks, flowing robes & fraternity paddle*" (O1).

Users especially challenged what they perceived as Yelp hiding behind the word "algorithm" when questioned about filtered reviews: "*Did we ever get an answer as to why his reviews are filtered? is 'Algorithm' the answer? the FAQ says 'Algorithm'*"' (O21). The opacity in operation was particularly frustrating for users who contacted Yelp to ask about their filtered reviews and received unsatisfying answers: "*Emailing yelp goes nowhere - just a form letter saying the same thing about how they don't want to give away how their formula works, but it 'considers many variables'*"' (O180). This lack of specificity left users without an understandable reason for being filtered, leading some to consider it "*a little underhanded that some of [the reviews] are hidden for unknown reasons*" (P11). Some even called Yelp "*a useless, fraudulent system*" (O57) as a result.

Demanding Transparency in Operation: As a result of the opacity in operation, users asked Yelp to "*unbox the algorithm*" (P3). They argued that if Yelp has "*some standard of what [is being filtered], they have to communicate that standard to people [...] so users know how they can avoid being filtered out because people want their review to count*" (P6). Even users who did not oppose the algorithm argued that "*it would be nice if Yelp was more up front about this [because] there would be far less disgruntled users questioning their algorithm*" (O32).

Some suggested that more transparency about the algorithm's operation might not require publishing standards or revealing how particular decisions are made. Instead, they would be satisfied with overall statistics about filtered and unfiltered reviews: "*I would love to see a global query in all posts within Yelp showing a graph of positive vs negative reviews, and would love to see a graph on the millions of post simple filtered*" (O164) though they then revealed their critical stance: "*then I would love to see how many post[s] were bogus with no cause or reason*" (O164).

Defending the Opacity in the Algorithm's Operation. The opacity in the algorithm's operation was not always seen as a drawback by the discussants. A few users (n_o=4) appreciated the opacity, expressing similar concerns to Yelp itself. Arguing, for example, that if the algorithm's operation was open to public, "*it would allow unscrupulous players to game the system*" (O18). To defend the algorithm's opacity in its operation, users also shared their own understandings of the "Catch-22" argument Yelp discusses on its blog [45]: "*If Yelp publicized the factors, then people would easily be able to 'fool' the filter, and the site would be more prone to spam, fake reviews, etc. It would be like putting a security system on your house and then posting the PIN on the front door*" (O44).

Investigating the Algorithm's Operation: Folk Theories. While discussing the opacity in operation, many users engaged in sense-making practices to investigate the algorithm's operation (n=87). These investigations resulted in a series of "folk theories": informal theories that users develop to perceive and explain how a system works. These theories, while non-authoritative, informal and sometimes incorrect, play an important role in how users interact with a system since they guide users' reactions and behavior towards a system [8, 12, 31]. Indeed, we find that users suggest very direct actions to adapt to the theories and ensure that their reviews are seen. In this section, we analyze users' theories about how the Yelp filtering algorithm works and how these theories can impact their attitude towards the system.

The Reviewer Engagement Theory: The most common theory of how the filtering algorithm works was based on the level of a reviewers' engagement with Yelp. Many users (n=66) believed that "*the activity is the main thing that keeps you in circulation*" (O98). Therefore, "*the more you participate in the site, the more Yelp will take into consideration that you an actual human*" (O32), and "*the more likely your reviews will show up!*" (O26). Users suggested many different factors that affect a user's engagement level with the site:

- *Reviewing Experience*: The main engagement factor users suggested was the amount of experience the reviewer had, as measured by the age of their account (n=9) and the number of reviews the user had posted on Yelp (n=41): "*The reviewer's Yelp experience is definitely a factor. If the reviewer is new to Yelp, has [...] few or no prior reviews, that reviewer's credibility scores much lower than someone already having a history of reviews over a period of time*" (O48). Therefore, a suggestion to users with filtered reviews was "*to write some more reviews & keep on Yelpin' if you'd like them to come out of the filter*" (O187).
- *Friendship Network*: Users (n=26) also theorized that the number of friends a user has will influence the filtering algorithm: "*you're sending up flags by [...] not having friends*" (O51). To address this, users suggested that "*a few friends and a few more reviews will take you out of the 'filter algorithm' and allow your reviews to be posted*" (O63).
- *Profile Completeness*: Many users (n=24) stated that "*Yelp requires that you completely fill out your profile*" (O32), and suggested adding a profile picture, because accounts with "*no profile picture [...] the algorithm will filter them out because they look like fake profiles*" (O145). Some even offered more specific suggestions to adapt to the algorithm: "*I definitely agree about a clear pic of yourself, the algorithm looks for things like that*" (O206).
- *Providing or Receiving Feedback*: Another common factor users (n=14) proposed was providing or receiving feedback for reviews: "*if your review earns FUC's (Funny, Useful, Cool), it increases the likelihood of it being recommended*" (O162) because "*while the algorithm is certainly secret, its a pretty safe bet that the rest of the community indicating they found value in the review could help it make it back to the front page*" (O79). Thus, the reviewers' folk theories of the algorithms operation could drive community engagement, as in one user who wrote, "*You should also be sure to FUC his review if you feel that way about it*" (O79).

The Review Format Theory: The other main theory about how the algorithm filters a review revolved around the review itself. Users (n=37) described different features of a review that can impact the algorithm's decision to filter it:

- *Extreme Language/Rating*: Some (n=24) believed that "*fake reviewers tend to have the BEST or the WORST experiences*" (O45). Therefore, "*if the review seems exceptionally over the top giddy about how great your business is, or exceptionally negative, [the algorithm] may also pick those reviews out for filtering as well, thinking the great ones may have been submitted by someone affiliated with the business, or the poor ones may have been submitted by a competitor*" (O151). So, users suggested those who had filtered reviews "*avoid liking everything or hating everything*" (P4).
- *Details*: A lack of details in a review was another reason some users (n=15) suggested a review might get filtered: "*Writing 2 sentences and giving a sterling or abysmal rating is a sure way to get that review flagged and voted as suspicious*" (O100). Therefore, they suggested users with filtered reviews to "*be more descriptive*" (O67). Indeed, some suggested that reviewers go beyond the review text and also "*add pictures*" (P9) to their reviews.

- *Length*: Distinct from adding details, some (n=11) simply suggested writing "*lengthy*" (P12) reviews to avoid filtering.

The above theories, while not necessary validated, still impacted how users interact with the system. Users tried to use their theories to remove their reviews from the filter. Some actions they could take themselves – adding more details, pictures or more reviews – but for others they called on the Yelp community for help – asking for votes on their reviews. Later in the paper, we discuss how users also used the theories they developed to write for the algorithm when we added transparency about it using our tool.

Questioning the Algorithm's Operation. While developing their theories about how the algorithm works, many users (n=60: n_o=50 & n_p=10) questioned the algorithm's operation when they found it "*inconsistent*" (P11) or "*arbitrary*" (O210). Users struggled because every theory was faced with exceptions. As some wrote, they "*can't imagine what the rationale that the algorithm is using could possibly be*" (O180) because "*everything [they] try to figure out was in common with many [filtered reviews], there were a bunch that didn't fit that pattern*" (P14). And similarly, when many users turned to common theories about how the algorithm works to get their own reviews unfiltered, those theories did not work: "*I even included photos from my meal [in the review]! What else can I provide to prove I was there? DNA?*" (O34).

This perceived lack of consistency in the algorithm's operation, along with the lack of response from Yelp, caused confusion and even resignation among users. Some even felt that they did not have a right to inquire why their reviews were filtered: "*But maybe this is all like Biology; you cannot really ask why, it just is*" (O54). This was particularly problematic for small business owners whose business's positive reviews were mostly filtered by the algorithm, since they could not even choose to remove their business page from Yelp: "*The filter needs to be modified but it won't. [...] Can't opt out, can't delete...so what do we do if we can't get out?*" (O102).

(Advertising) Bias: Users' perceived lack of algorithmic agency, together with the inconsistency of the algorithm's operation, caused fears that Yelp biases results to promote its own advertising business: "*you pay Yelp you get more positive [reviews] and they hide your negative [reviews]*" (O144). Many business owners claimed they received calls from the Yelp Sales team about advertising with Yelp, and as soon as they rejected the offer, positive reviews began being filtered: "*a bit coincidental when examining the timeline vs conversations with the Yelp Sales Team*" (O102). These business owners considered this "*a common tactic used by Yelp to 'strong arm' small businesses to advertise with Yelp*" (O135) which "*SMELLS LIKE EXTORTION!*" (O144). This alleged practice has been the subject of multiple lawsuits [1, 26].

This claim, regardless of its validity, drove many users to argue that "*Yelp has finally crossed over from an honest user generated content review site, to a capitalist advertisement platform*" (O121). This opinion was held not only by business owners, but also by reviewers. When they could not find a pattern in how reviews were filtered, some participants stated that "*maybe the automated software is a lie [and] they can take out the reviews they don't want*" (P2). This show how a lack of transparency, combined with a perceived inconsistency in an algorithm's operation, can lead users to suspect bias.

Demanding a Change in Operation: Users (n=16) argued that Yelp shouldn't use "the algorithm" as an excuse for deceptive practices: "*Oh, my toilet, watch, car and email is 'automated' too...but doesn't mean I can change or improve the 'automation'. Stop hiding behind that please*" (O102). These users even proposed improvements they wanted to see:

- *Adding Human in the Loop*: Some users (n=7) suggested Yelp "*add a human touch to the algorithm, or they're gonna continue to get allegations that Yelp reviews are paid for*" (O163). They also suggested an appeal process to include human oversight: "*My only wish is that it had a button on the filtered reviews for people to click on, so if an Elite member or regular user of this site felt that one or more of the filtered reviews was actually genuine, they could click on that and send it to a human for their review*" (O151).
- *Adding or Modifying a Variable*: Some users (n=8) suggested adding or changing variables. For example, users suggested that "*every review that is posted on a business should affect their overall star rating*"(P8) even if it is filtered. Others suggested adding variables like a user's account age or IP (though some believed that these variables are already included in the algorithm).
- *Equality vs Equity*: A few users (n=2) argued that the algorithm may have disparate, unequal impacts on different communities, and asked for a change in the algorithm's operation to provide equity, rather than equality. That is, while it may seem fair that the algorithm treats all businesses the same way, some communities may be more harmed by the algorithm's impacts. For example, if a business can only do three jobs per month (e.g., a sculptor), but their reviews are filtered and their score is calculated in the same way as busy restaurants, they may suffer lower overall ratings: "*[the] problem is that the quantity and type of my customers doesn't fit within Yelp's one-size-fits-all approach. I've struggled to get 3 reviews, and guess what? All of them [are] "not recommended" but are all legitimate clients and reviews. I sincerely hope that Yelp changes their algorithm to be more industry specific [... and help] grow small business with few but precious clients*." (O193)

Defending the Algorithm's Operation. In our analysis, we found some users (n_o=23) who defended the algorithm's operation, arguing that "*while 'the algorithm' isn't perfect and isn't always 100% accurate, in most cases it does its task and does a decent job of keeping the reviews honest*" (O151). They also argued that the algorithm has the right to be imperfect in operation because it is faced with so many fake reviews: "*the Yelp review filter is like the liver. It flushes out toxins. Sometimes it doesn't work because of excessive drinking*" (O117). These users suggested that those complaining about the algorithm's operation "*give it time [and] keep reviewing*" (O154) until they "*gain the 'bot respect'*" (O82).

The Algorithm is Not Biased towards Advertising: In addition to defending the algorithm's imperfection, some users (n=10) argued that "*advertising has nothing (zero, zilch, nada) to do with [...] the way the review filter works*" (O104). These users often made reference to the fact that "*not one lawsuit [about Yelp's advertising bias] has succeeded and the biggest class-action one was thrown out of court*" (O105). Some also argued that "*there are plenty of successful businesses on Yelp that don't advertise*" (O138), adding the fact that "*there are examples all over Yelp of advertisers that have positive reviews filtered and non-advertisers with negative reviews filtered*" (O18). Using these arguments, users tried to defend the algorithm and convince others that while the algorithm's operation is imperfect, it is not biased towards advertising.

Factors Associated with Different Perceptions (RQ2)

We found that while some users challenged the algorithm and its opacity, many others defended it. But why do some users challenge the algorithm while others passionately defend it? Past work has not yet captured what factors are associated with different perceptions of and stances towards an algorithm. We hypothesize that these users' interactions with the system may influence their stance.

To study this, we first devised a coding scheme to capture the stance of a discussant/participant towards the algorithm. This coding scheme labels a user as i) an algorithm "*challenger*" if she only challenged the algorithm, ii) an algorithm "*defender*" if she only defended the algorithm, or iii) a "*hybrid*" if she challenged the algorithm in one aspect but defended it in another. For example, a participant would be labeled a hybrid if she argued against the opacity of the algorithm's existence but defended the algorithm's operation overall. Using this scheme, we coded 150 Yelp users' stance towards the algorithm ($n_{challenger}$=69, $n_{defender}$=61, n_{hybrid}=20).

During the coding process, we observed that users who were more engaged with the Yelp platform (e.g. wrote many reviews, provided or received feedback on reviews, or had a large number of friends) usually defended the algorithm and/or its opacity. We also found that elite users usually strongly defended the algorithm, while business owners questioned it. To gain a more precise understanding, we ran a statistical analysis to find the influence of the *engagement level* of a user with the platform as well as the users' *personal gain* from the system on their stance with respect to the algorithm. This analysis shows what correlates with users' stance on the algorithm, but not necessarily what causes this stance.

Engagement level. To capture the engagement level of a user, we selected features from the Yelp user profile. Features were selected on the basis of the "reviewer engagement theory" captured in RQ1. We collected 13 engagement features (such as number of reviews, friends, and compliments) for each user. We then ran a pairwise correlation analysis on these features and the stance of a user toward the algorithm. As Table 1 shows, the more engaged a user is with Yelp, the more she/he defends the algorithm in its existence, operation, and/or opacity. This may be because as a user engages with an algorithmic platform, they begin to understand the platform's dynamics better. Greater engagement provides users an opportunity to "play" with the algorithm and investigate its inputs and outputs. Prior work has shown that users who actively engage with an algorithm are more likely to be aware of its existence [14]. Our study, however, goes further, showing that user engagement is associated not only with a users' awareness of the algorithm, but also with the user's stance towards the algorithm.

Personal Gain. We also studied the algorithm's impact on a user's life, that is, their personal gain from the system. In particular, we measured these in the form of user types: elite users and business owners. We collected these either from their profile (if they were elite, and if so, for how many years) or from their discussion (if they asserted that they own a business page on Yelp). The same statistical analysis as above revealed that being an elite user positively correlates with defending the algorithm, while being a business owner negatively correlates with defending the algorithm.

To understand the underlining reasons, we investigated the discussions of elite users and business owners from the online dataset. Many users argued that for elite users, the algorithm has a positive *reputational* and *social* impact on their online status. First, the algorithm has a role in deciding who is an elite member and once an elite members, users' reviews are rarely filtered. Furthermore, Yelp provides elite users with opportunities like social gatherings with other elite members, free food and beverage, and the authority to act as a Yelp Ambassador (e.g., answering other users' questions). Such privileges led some elite users to call themselves a part of Yelp, using phrases like "we" or "us" when addressing other users' complaints about the algorithm.

Business owners, on the other hand, argued that the algorithm has a negative impact on their *financial* status. They

Variable	Coefficient	p-value
Engagement level		
Review count	0.460	0.00**
Friend count	0.369	0.00**
First review count	0.356	0.00**
Update count	0.346	0.00**
Compliment count	0.339	0.00**
Feedback count (Funny,Useful,Cool)	0.329	0.00**
Account age	0.318	0.00**
Lists count	0.308	0.00**
Tips count	0.270	0.00**
Follower count	0.265	0.00*
Has photo?	0.243	0.00*
Bookmark count	0.221	0.00*
Photo count	0.138	0.09
Personal gain		
Elite member?	0.586	0.00**
Business owner?	-0.227	0.00*

** $p < 0.001$ * $p < 0.01$

Table 2: The correlation between the engagement level as well as personal gain and user's stance towards the algorithm.

stated that the review filtering process caused "*a loss of income to [their] business on some fronts*" (O102) since it moved some of their positive reviews to the filter.

Adding Transparency about the Algorithm (RQ3)

We found that only three of the 15 interviewees were aware of the Yelp review filtering practice. The others stated that they "*had no idea that that was a thing*" (P11) which made them "*aggravated and annoyed*" (P3). Some who saw their filtered review at the top of a business's page thought that "*maybe because it's such a good review that maybe they put it at the top*" (P12) but found moments later that it was filtered.

We also found this lack of awareness among discussants (n_o = 13): "*Little did I know my independent reviews posted in earnest with no ties to any business whatsoever were considered suspect by the mysterious Yelp algorithms*" (O1). The fraction of unaware users was much lower among the discussants, however, most likely because participating in an online discussion about the algorithm would either require or prompt knowledge of the algorithm's existence.

Change of Attitude. After being exposed to the algorithm and their filtered reviews via *ReVeal*, participants who were unaware of the filtering process expressed their intent to change their behavior.

Writing for the Algorithm: Six participants stated that they will change the way they were writing their reviews: "*I would write for the algorithm as my audience [...] to get my reviews unfiltered* (P3). They said that "*they would put a little more thought into*" (P12) writing a review, and in doing so, they used the theories they developed during exploring filtered and unfiltered reviews on Yelp in the Interview. For example, some said that they "*will probably add pictures (to their reviews) because [they] noticed that the filtered ones didn't have pictures*" (P9). "*Making [their reviews] lengthy*" (P12) and "*doing a review that has more detail*" (P4). These results add practical evidence to Gallagher's discussion about "algorithmic audience" [18] and corroborate previous findings about transferring the knowledge users gain through an algorithm visualization tool to their behavior [14].

Using the System Less or Leaving It: About half of the participants (n_p = 7), however, stated that they "*won't probably use Yelp that much*" (P1) or they "*would probably stop using Yelp*" (P3) "*because if [your review] is just going to be grayed out in the bottom and no one's ever going to see it, why are you going to put the total of the time and effort for it?*" (P15). Therefore, some added that they "*want to use a different [reviewing platform]– like TripAdvisor or Google*" (P2). We found a similar pattern among some discussants (n_o = 10) as well. The statement of leaving an algorithmic platform due to the presences of an opaque and biased algorithm has been found in previous work as well [15].

5 LIMITATIONS AND ETHICAL CONSIDERATIONS

In developing *ReVeal*, we attempted to use the Yelp API to comply with its terms of service. However, because the Yelp API does not allow access to filtered reviews, we collected this information via scraping. Nevertheless, we avoided accessing any users' private data; all data *ReVeal* used for analysis was public data available to any online user. However, we note that while the data in the online dataset was public, researchers have noted potential issues with public data, for example, privacy issues and dealing with future deletions [30]. Therefore, we have excluded from analysis any discussion touching on topics that might be considered private. Finally, when adding transparency about the algorithm, we studied self-reported intentions, but have no data on future usage of the system. A longitudinal study should investigate if users' usage behavior matched their reported intentions.

6 DISCUSSION & FUTURE WORK

Our findings reveal that adding transparency to an opaque and potentially biased algorithmic system can provide users with a more informed interaction with the system. Here, we discuss a set of design implications about adding different levels of transparency to opaque algorithmic processes. We also propose changes in the design of algorithmic platforms to empower users as auditors.

Transparency: A Solution or A Challenge?

While our work shows that adding transparency into an opaque algorithm can benefit users via a more informed interaction with the system, it is not clear how much transparency is enough. Here, we discuss transparency in algorithmic existence, operation, and also impact.

Transparency in Existence. Our results found that while users defended the algorithm's existence, operation, and even its opacity in operation, no user defended the opacity of the very *existence* of the algorithm (Table 1). That is, while users may see the algorithm as supporting the functioning of the platform, they want to know that it exists. Opacity in the algorithm's existence can be considered a deception, resulting in a breakdown of trust between users and the system.

Designing a system to signal an algorithm's existence, however, is not straightforward. For example, previous work has shown that while Facebook has provided two options ("top stories" and "most recent") in its News Feed that imply the presence of a feed curation algorithm, many users were still not aware of this algorithm [14]. *ReVeal*'s design suggests the presence of a filtering algorithm by changing the color of the filtered reviews. However, whether this design is explicit enough needs more investigation. Another challenge in revealing the presence of a filtering algorithm is user discouragement. Seeing that their reviews have been filtered can make some Yelp users frustrated, even stating they plan to leave the platform. Therefore, in addition to signaling an algorithm's existence, designs may need to signal the necessity of the algorithm as well.

Transparency in Operation. Previous work [12], as well as our results, suggest that users who build theories about an algorithm's operation act on those theories. Therefore, the level of transparency a system provides about an algorithm's operation affects users' usage behavior. However, given the complicated internal process of an algorithm, it is usually impossible to make an algorithm's operation fully transparent via design. Such design would also likely complicate users' interactions with the system. Another issue in making an algorithm's operation transparent is the potential for malicious users gaming the system.

However, the fact that many algorithmic systems (such as Yelp's review filter) are housed in black boxes poses a substantial challenge to outside researchers trying to evaluate different levels of transparency. While past research has attempted to reverse engineer the Yelp filtering algorithm [25], no ground truth is available to those outside Yelp.

Beyond Existence and Operation: Transparency in Impact. While our work focused on the existence and operation of the Yelp review filtering algorithm, our results also suggest a third important factor: the algorithm's impact. Our work studied the impacts of users' perceptions of the existence and operation of Yelp's opaque review filtering algorithm. However, we found that even when people have a sense of the existence and operation of the algorithm, they often do not have a sense of how the algorithm impacts them.

As our results showed, when business owners do not understand how the algorithm affects their business' score, they can fear that it is lowering scores, or even that Yelp intentionally lowers scores to force them to buy advertising. However, past work has shown that there is no significant difference in the ratings of reviews that are filtered between businesses that pay for advertising versus those that do not [29].

We believe adding transparency into algorithmic impacts could help address these user concerns. Given that Yelp's review filtering algorithm does not appear to be biased, the interface design should showcase their efforts to keep the system fair. For example, the interface might surface this one aspect of the algorithm's impacts, showing business owners their final rating both with and without the filtered reviews. Adding the impacts of an algorithm can be adopted in other opaque algorithmic platforms, particularly when the reality of the impact of an algorithm differs from users' perceptions.

Users as Auditors

Amidst the numerous proposals to better understand and investigate opaque algorithmic systems [9, 19, 33, 38], one thrust has focused on auditing these systems. Sandvig et al. proposed a taxonomy of auditing approaches [37], from studying the code directly to collaborative audits, but all require the intervention of researchers, regulators or other third parties to coordinate. Our results highlight a new form of work, with users discussing, questioning, and defending the review filtering algorithm themselves on Yelp's forums. This is a new form of audit: a *collective audit*, driven purely by users in a collective attempt to understand how an algorithm works.

Providing an Auditing Platform from Within. Whether intentional or not, Yelp has helped support this collective audit by its users through its interface design. Past work on watchdogs from within [15] addressed only improvisational uses of the system (for example, integrating comments about bias into their review text or changing how they score reviews). As we have shown in this work, the Yelp interface provides a space to take this user audit further. By providing an online discussion forum – a single, focused location where users can discuss with others – the Yelp platform supports the kinds of awareness raising identified in prior work, but also a deeper engagement with the algorithmic operation, including defense of the algorithmic design that we uncovered in our work.

The discussion forums provide a stronger version of the watchdog from within in three ways: 1) the forums are integrated within the platform itself, and since users may perceive their discussions as being permitted by the platform, they may feel greater agency and support, 2) the discussion is within the platform, rather than external so more users are likely to discover it, and 3) it is a discussion forum, so users engage with each other – people get feedback on their ideas, learn from each other, develop and even change their thinking. Unlike comments integrated into reviews, this style of forum has the potential to help users adapt and grow in their understanding and perception of the algorithm. Developers can leverage similar on-platform community forums on other algorithmic systems to solicit input on algorithms from a variety of users and create communities that foster algorithmic agency.

Algorithm Bug Bounty. While on-platform communities can provide users with an opportunity to discuss possible algorithmic biases, we believe designing algorithmic platforms so that users can report bias directly can provide users with even more algorithmic agency. This reporting process can also be incentivized, analogous to security's "bug bounty" programs [11] in which companies incentivize users to conduct security research and report flaws for monetary and reputational gain while providing legal protection from the applicable anti-hacking laws. Embedding such design practices in opaque algorithmic systems not only can empower users, but also can increase users' trust in the system. As our results showed, users may identify unexpected aspects of a system as biased (e.g., disparate impacts on smaller businesses) and thus these "bias bugs" may also help system designers better understand their users' needs.

7 CONCLUSION

We found that users challenge or defend a potentially biased algorithm in its existence, operation, and opacity; and their stance depends on how engaged they are with, and how much personal gain they get from the algorithm. Adding transparency into the algorithm, however, can change users' attitudes towards an algorithm. These findings uncover many opportunities and challenges in designing the opaque algorithmic systems that might be biased – or might simply be perceived as biased. We argue that as more algorithmic systems exert power to shape users' experiences, system designers need to communicate the existence, operation, and the impact of opaque algorithmic processes. However, finding the right level of transparency to avoid complicating users' interactions or providing opportunities for gaming is a challenging task which needs a careful design process. We hope our findings can drive future research into the design of opaque algorithmic systems, particularly when their opacity can cause concerns over potential bias or deception.

8 ACKNOWLEDGMENTS

This work was funded in part by Google and NSF grant CHS-1564041.

REFERENCES

[1] 2014. Levitt v. Yelp, 765 F.3d 1123 (9th Cir. 2014). https://caselaw.findlaw.com/us-9th-circuit/1676994.html.

[2] Mike Ananny and Kate Crawford. 2018. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society* 20, 3 (2018), 973–989.

[3] Michael Anderson and Jeremy Magruder. 2012. Learning from the crowd: Regression discontinuity estimates of the effects of an online review database. *The Economic Journal* 122, 563 (2012), 957–989.

[4] Athanasios Andreou, Giridhari Venkatadri, Oana Goga, Krishna P Gummadi, Patrick Loiseau, and Alan Mislove. 2018. Investigating Ad Transparency Mechanisms in Social Media: A Case Study of Facebook's Explanations. In *The Network and Distributed System Security Symposium (NDSS)*.

[5] Jenna Burrell. 2016. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society* 3, 1 (2016), 2053951715622512.

[6] Matthew Chalmers and Ian MacColl. 2003. Seamful and seamless design in ubiquitous computing. In *Workshop at the crossroads: The interaction of HCI and systems issues in UbiComp*, Vol. 8.

[7] Matthew Crain. 2016. The limits of transparency: Data brokers and commodification. *new media & society* (2016).

[8] Michael A DeVito, Darren Gergle, and Jeremy Birnholtz. 2017. Algorithms ruin everything:# RIPTwitter, folk theories, and resistance to algorithmic change in social media. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 3163–3174.

[9] Nicholas Diakopoulos. 2014. Algorithmic-Accountability: the investigation of Black Boxes. *Tow Center for Digital Journalism* (2014).

[10] Nicholas Diakopoulos and Michael Koliska. 2017. Algorithmic transparency in the news media. *Digital Journalism* 5, 7 (2017), 809–828.

[11] Amit Elazari Bar On. 2018. Private Ordering Shaping Cybersecurity Policy: The Case of Bug Bounties. (2018). Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3161758.

[12] Motahhare Eslami, Karrie Karahalios, Christian Sandvig, Kristen Vaccaro, Aimee Rickman, Kevin Hamilton, and Alex Kirlik. 2016. First I "like" it, then I hide it: Folk Theories of Social Feeds. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2371–2382.

[13] Motahhare Eslami, Sneha R Krishna Kumaran, Christian Sandvig, and Karrie Karahalios. 2018. Communicating Algorithmic Process in Online Behavioral Advertising. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 432.

[14] Motahhare Eslami, Aimee Rickman, Kristen Vaccaro, Amirhossein Aleyasen, Andy Vuong, Karrie Karahalios, Kevin Hamilton, and Christian Sandvig. 2015. I always assumed that I wasn't really that close to [her]: Reasoning about Invisible Algorithms in News Feeds. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 153–162.

[15] Motahhare Eslami, Kristen Vaccaro, Karrie Karahalios, and Kevin Hamilton. 2017. "Be Careful; Things Can Be Worse than They Appear": Understanding Biased Algorithms and Users' Behavior Around Them in Rating Platforms.. In *ICWSM*. 62–71.

[16] Megan French and Jeff Hancock. 2017. What's the folk theory? Reasoning about cyber-social systems. (2017). Available at SSRN: https://ssrn.com/abstract=2910571.

[17] David Adam Friedman. 2017. Do We Need Help Using Yelp: Regulating Advertising on Mediated Reputation Systems. *U. Mich. JL Reform* 51 (2017), 97.

[18] John R Gallagher. 2017. Writing for algorithmic audiences. *Computers and Composition* 45 (2017), 25–35.

[19] Tarleton Gillespie. 2014. The relevance of algorithms. *Media technologies: Essays on communication, materiality, and society* 167 (2014).

[20] Eric Goldman. 2014. Court Says Yelp Doesn't Extort Businesses. *Forbes,*https://www.forbes.com/sites/ericgoldman/2014/09/03/court-says-yelp-doesnt-extort-businesses/#72d067ad6e4a.

[21] Bryce Goodman and Seth Flaxman. 2016. EU regulations on algorithmic decision-making and a "right to explanation". In *ICML Workshop on Human Interpretability in Machine Learning*.

[22] Jim Handy. 2012. Think Yelp is Unbiased? Think Again!! *Forbes,*https://www.forbes.com/sites/jimhandy/2012/08/16/think-yelp-is-unbiased-think-again/#76ddbc6811d1.

[23] Jonathan L Herlocker, Joseph A Konstan, and John Riedl. 2000. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*. ACM, 241–250.

[24] Shagun Jhaver, Yoni Karpfen, and Judd Antin. 2018. Algorithmic Anxiety and Coping Strategies of Airbnb Hosts. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 421.

[25] David Kamerer. 2014. Understanding the Yelp review filter: An exploratory study. *First Monday* 19, 9 (2014).

[26] Inkoo Kang. 2013. Read nearly 700 FTC complaints regarding Yelp "Please help the business people from the Internet Mafia". https://www.muckrock.com/news/archives/2013/jan/23/businesses-yelp-thug-of-the-internet/.

[27] René F Kizilcec. 2016. How much information?: Effects of transparency on trust in an algorithmic interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2390–2395.

[28] Min Kyung Lee, Daniel Kusbit, Evan Metsky, and Laura Dabbish. 2015. Working with machines: The impact of algorithmic and data-driven management on human workers. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 1603–1612.

[29] Michael Luca and Georgios Zervas. 2016. Fake it till you make it: Reputation, competition, and Yelp review fraud. *Management Science* 62, 12 (2016), 3412–3427.

[30] Jim Maddock, Kate Starbird, and Robert M Mason. 2015. Using historical twitter data for research: Ethical challenges of tweet deletions. In *CSCW 2015 Workshop on Ethics for Studying Sociotechnical Systems in a Big Data World. ACM*.

[31] Don Norman. 2013. *The design of everyday things: Revised and expanded edition*. Constellation.

[32] Nvivo 2017. Nvivo 11. http://www.qsrinternational.com/nvivo-product.

[33] Frank Pasquale. 2015. *The black box society: The secret algorithms that control money and information*. Harvard University Press.

[34] Pearl Pu and Li Chen. 2007. Trust-inspiring explanation interfaces for recommender systems. *Knowledge-Based Systems* 20, 6 (2007), 542–556.

[35] Emilee Rader, Kelley Cotter, and Janghee Cho. 2018. Explanations as Mechanisms for Supporting Algorithmic Transparency. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 103.

[36] 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*. ACM, 173–182.

[37] Christian Sandvig, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. 2014. Auditing algorithms: Research methods for detecting discrimination on internet platforms. *Data and discrimination: converting critical concerns into productive inquiry* (2014).

[38] Nick Seaver. 2013. Knowing algorithms. *Media in Transition* 8 (2013), 1–12.

[39] Andrew D Selbst and Solon Barocas. 2018. The intuitive appeal of explainable machines. (2018).

[40] Rashmi Sinha and Kirsten Swearingen. 2002. The role of transparency in recommender systems. In *CHI'02 extended abstracts on Human factors in computing systems*. ACM, 830–831.

[41] Harry Tucker. 2016. Australian Uber drivers say the company is manipulating their ratings to boost its fees. *Businessinsider*.

[42] Weiquan Wang and Izak Benbasat. 2007. Recommendation agents for electronic commerce: Effects of explanation facilities on trusting beliefs. *Journal of Management Information Systems* 23, 4 (2007), 217–246.

[43] WhiteHouse. 2016. Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights. *Washington, DC: Executive Office of the President, White House* (2016).

[44] Yelp. [n. d.]. Yelp Does Not Extort Local Businesses or Manipulate Ratings. https://www.yelp.com/extortion.

[45] Yelp. [n. d.]. Yelp's Recommendation Software Explained. https://www.yelpblog.com/2010/03/yelp-review-filter-explained.