
Information Systems Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

# Consequences of Information Feed Integration on User Engagement and Contribution: A Natural Experiment in an Online Knowledge-Sharing Community

Zike Cao, Yingpeng Zhu, Gen Li, Liangfei Qiu

To cite this article:
Zike Cao, Yingpeng Zhu, Gen Li, Liangfei Qiu (2023) Consequences of Information Feed Integration on User Engagement and Contribution: A Natural Experiment in an Online Knowledge-Sharing Community. Information Systems Research

Published online in Articles in Advance 14 Sep 2023

. https://doi.org/10.1287/isre.2022.0043

informs.
https://pubsonline.informs.org/journal/isre

INFORMATION SYSTEMS RESEARCH
*Articles in Advance*, pp. 1–23
ISSN 1047-7047 (print), ISSN 1526-5536 (online)

# Consequences of Information Feed Integration on User Engagement and Contribution: A Natural Experiment in an Online Knowledge-Sharing Community

**Zike Cao,[a] Yingpeng Zhu,[b] Gen Li,[c,]* Liangfei Qiu[d]**

[a] Department of Data Science and Engineering Management, School of Management, Zhejiang University, Hangzhou 310058, China; [b] Department of Accounting and Information Management, Faculty of Business Administration, University of Macau, Taipa, Macau, China; [c] Department of Information Management and Business Intelligence, School of Management, Fudan University, Shanghai 200433, China; [d] Department of Information Systems and Operations Management, Warrington College of Business, University of Florida, Gainesville, Florida 32611

*Corresponding author

**Contact:** caozike@zju.edu.cn, https://orcid.org/0000-0002-5378-4772 (ZC); yingpengzhu@um.edu.mo, https://orcid.org/0000-0002-1440-5100 (YZ); gli@fudan.edu.cn, https://orcid.org/0000-0001-9493-3976 (GL); liangfei.qiu@warrington.ufl.edu, https://orcid.org/0000-0002-8771-9389 (LQ)

**Received:** January 18, 2022
**Revised:** August 26, 2022; April 13, 2023
**Accepted:** July 31, 2023
**Published Online in Articles in Advance:** September 14, 2023

**https://doi.org/10.1287/isre.2022.0043**



**Abstract.** Many online communities that rely on effortful, voluntary content contributions offer additional content curation tools to facilitate social interactions and encourage user contributions. Any platform that offers two or more heterogeneous content types (e.g., expert knowledge and social posts) faces a choice about the presentation format: whether to display the content types separately or in an integrated information feed. We leverage a natural experiment on Zhihu, a Q&A platform that offers a social-interaction-oriented functionality called *Ideas*. Zhihu initially presented *answers* (expert knowledge content) and *ideas* (social posts) in two different information feeds, but the platform integrated *ideas* into the same information feed as *answers* in June 2019. We find that information feed integration significantly decreased user engagement with and contribution of both *ideas* and *answers*. We hypothesize that users decreased their engagement because the juxtaposition of incongruous types of content increased mindset switching and cognitive strain. This hypothesis is supported by an additional laboratory experiment. We also present evidence showing that contributions decreased both because of the decrease in engagement (weaker social recognition incentives) and because integration heightened concerns that posting *ideas* would dilute the contributor's professional image. Our findings have important theoretical and practical implications for any platform that hosts heterogeneous content.

**History:** Xiaoquan (Michael) Zhang served as the senior editor and Yili (Kevin) Hong served as associate editor for this article.

**Funding:** Z. Cao acknowledges this research was funded by National Natural Science Foundation of China [Grants 72201238, 72192823]. G. Li acknowledges this research was funded by National Natural Science Foundation of China [Grant 72102047] and Shanghai Pujiang Program [Grant 21PJC006].

**Supplemental Material:** The e-companion is available at https://doi.org/10.1287/isre.2022.0043.

**Keywords:** information feed integration • online Q&A communities • user-generated content (UGC) • image concern • natural experiment • regression discontinuity in time (RDiT)

*Consumers use multiple self-presentation strategies to construct digital collages that represent the self. As one aspect of self is explored (e.g., professional, hobbyist), consumers are often motivated to use the medium to explore and display other selves* (Schau and Gilly 2003, p. 390).

## 1. Introduction

Many online communities thrive on user-generated content (UGC) that requires professional expertise and knowledge. Prominent examples include online encyclopedias (*Wikipedia*, *How Stuff Works*), open-source software communities (*Linux*, *Android*), and knowledge-sharing communities (*Quora*, *Stack Overflow*). To encourage users to contribute content, online communities may offer additional virtual spaces, so-called informal "third places," to facilitate social interaction among users (Chen et al. 2021). For example, Wikipedia offers *Talk* pages, where users can converse while collaborating on articles. Stack Overflow has a *Chat* function for informal communications, separate from the sharing of programming knowledge. The popular online discussion forum Reddit also offers a group chat feature. Quora allows users to contribute free-form content in a social networking feature called *Posts*.[1]

Knowledge-sharing platforms may choose to offer social-interaction-oriented functionalities for multiple reasons. For one, social benefits and recognition are important drivers of voluntary contributions online (e.g., Zhang and Zhu 2011; Goes et al. 2014, 2016), so additional opportunities for social interactions may increase users' incentives to contribute. For another, users may feel compelled to display multiple selves online (Schau and Gilly 2003), and a social-interaction-oriented functionality enables users to freely express themselves, as they could on social networking platforms such as Facebook or Twitter.

Any platform that offers two or more heterogeneous content types (e.g., expert knowledge and social posts) faces a choice about the presentation format: namely, whether to display the content types separately or in an integrated information feed. The information presentation format is known to affect behaviors and decisions in various disciplines, including information systems, accounting, finance, and marketing (Bettman and Kakkar 1977, Maines and McDaniel 2000, Kim and Dennis 2019). Yet, there is no consensus among platform managers about the best approach. For example, Quora displays *answers* to questions and social *posts* in an integrated information feed, while Sina Weibo displays long articles in a separate channel from the main information feed, which is dominated by short content. See Online Appendix A for screenshots of the interfaces.

We study Zhihu, the leading Chinese question-and-answer (Q&A) platform, which launched a social-interaction-oriented functionality called *Ideas* in August 2017.[2] Zhihu is known for the Q&A format, where users ask *questions* and other users (often those with professional, expert knowledge of the topic) post *answers*, while the new *Ideas* channel is designed to be a social networking feature. The two types of content, *answers* and *ideas*, differ on various dimensions, including format, contribution effort, contribution motivation, and intended purpose. Table 1 summarizes these differences.

As shown in Table 1, *answers* typically are structured content targeted at questions raised by other users in the community. Users post *answers* mainly to share their expert knowledge and help address questions other users have. By contrast, *ideas* are free-form content and do not need to pertain to pre-existing questions. Users post *ideas* mainly out of self-expression need to present multiple selves on the same platform (Schau and Gilly 2003). Therefore, users often use the *Ideas* channel to share their personal life (e.g., hobbies, leisure activities, etc.) or spontaneous thoughts. Users could also use the *Ideas* channel to post content on serious topics, but such content, compared with content in the format of *answers*, is shorter and less in-depth, and it is mainly for publicity purpose (e.g., expressing their interests in certain topics) instead of sharing their own expert knowledge to help answer others' questions.

As a result, *answers* tend to be longer and more in-depth than *ideas*, and *answers* (versus *ideas*) usually require higher effort to generate. In terms of the underlying contribution motivations, the two types of content are also fundamentally different: posting *answers* is mainly instrumental for building a professional reputation (Wasko and Faraj 2005, Lou et al. 2013), which can also translate into tangible financial rewards on Zhihu (Wang et al. 2022), while posting *ideas* is mainly due to expressive and social motives (Wu 2013, Chen et al. 2021).[3] On Quora, a platform similar to Zhihu, there also exist similar differences between an *answer* and a *post* (analogous to an *idea*) as shown in Figure 1.

Although Zhihu initially presented *answers* (expert knowledge content) and *ideas* (social posts) in two different information feeds, Zhihu abolished the separate information feed for *ideas* in June 2019. After that, all *ideas* were integrated into the same information feed as *answers*. We treat this sudden change as a natural experiment to study how the presentation format of heterogeneous types of UGC (separate or integrated information feeds) affects user engagement and contribution behaviors on a Q&A platform.

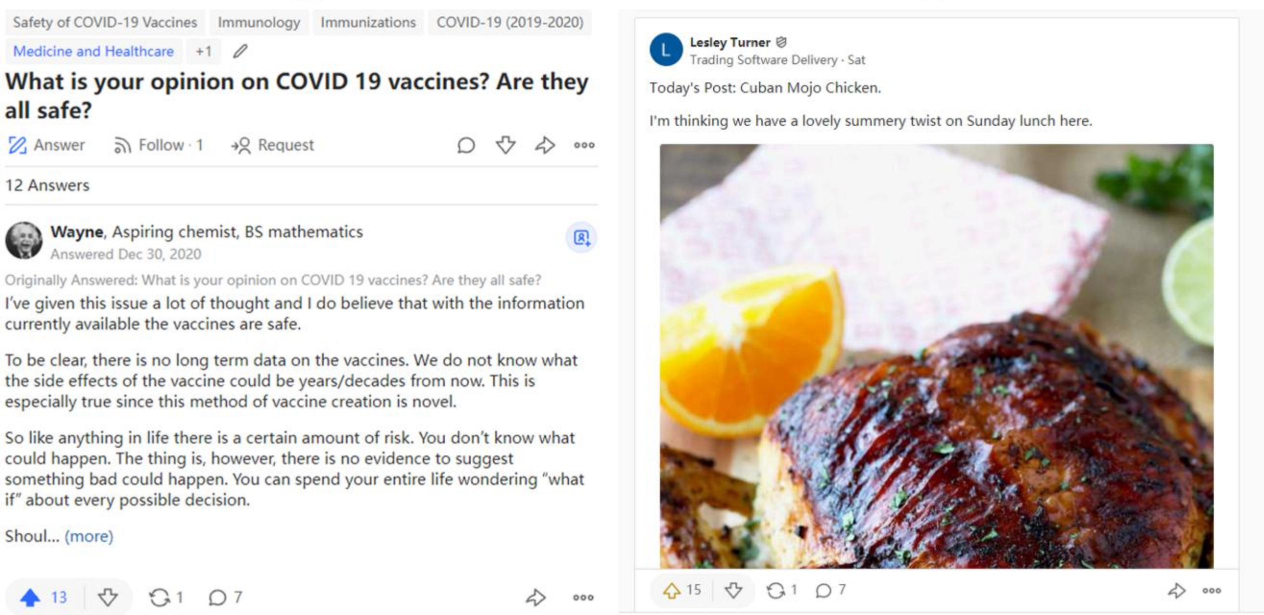
The integration of the information feeds for heterogeneous content (or "integration," for short) could plausibly increase or decrease user engagement with each type of content—both possibilities have some support in the literature. On Zhihu, integration increased the size of the audience that would potentially view both content types, and it increased the diversity and richness of information (Wu et al. 2019, Esfandiari et al. 2021) that was available for low search costs (Palmer 2002, Agnew and Szykman 2005). Thus, integration might have increased user satisfaction, which can increase engagement (Ma and Agarwal 2007). However, users likely need to use different "mindsets" to process *answers* and *ideas*, so integration might have increased cognitive strain and decreased information processing fluency due to increased demand on mindset switching (Meyers-Levy and Tybout 1989, Speier et al. 1999, Aggarwal and

**Table 1.** Summary of Differences between *Answers* and *Ideas* on Zhihu

| | *Answers* | *Ideas* |
|---|---|---|
| Format: | Structured; long; in-depth; targeted at existing questions | Free-form; short; brief; requiring no existing questions |
| Purpose & Motivation: | Knowledge sharing; professional reputation building | Self-expression; social interaction |
| Effort per content: | High | Low |

**Figure 1.** (Color online) Examples of User *Answers* and *Posts* on Quora

(a) (b)




*Notes.* (a) Answer. (b) Post.

McGill 2007, Hamilton et al. 2011, Yan et al. 2018). If so, then users may have consumed less content and engaged less to reduce the discomfort (Festinger 1957, Kruglanski et al. 2018).

The effects of integration on the incentives to contribute each type of content also are theoretically ambiguous. The effect of integration on user engagement, as described in the previous paragraph, may translate directly into an effect on contribution incentives. If integration led to more engagement, then users might have been more motivated to contribute both *answers* and *ideas* because they could anticipate more social recognition benefits from contributing (von Krogh et al. 2012; Chen et al. 2018, 2022; Kuang et al. 2019); the opposite would be expected if integration decreased engagement. Moreover, previous research suggests that social image is an important antecedent of UGC (e.g., Wasko and Faraj 2005, Ma and Agarwal 2007, Goes et al. 2016, Qiu and Kumar 2017, Chen et al. 2018, Pu et al. 2020), and the image of a "professional expert" usually is desirable in online knowledge-sharing communities (Wasko and Faraj 2005, Lou et al. 2013). The integration of expert knowledge content with social content may cause users to worry that posting the latter will dilute their professional expert image (Shanteau 1975, Nisbett et al. 1981), leading to a decrease in the posting of *ideas*.

In sum, the consequences of integrating the information feeds for heterogeneous content are theoretically unclear and, until now, empirically untested. Our results from the natural experiment on Zhihu show that, on average, information feed integration significantly *decreased* both user engagement with *ideas* and the contribution volume of *ideas*. The same effects seemed to also occur for *answers*, though at smaller magnitudes.

The negative effects on engagement are consistent with the hypothesis that the juxtaposition of heterogeneous UGC impedes information processing and acquisition due to the mindset switching effect. We provide additional evidence for the information processing mechanism with a randomized controlled laboratory experiment. We find that subjects in the Mixed condition (*answers* and *ideas* interspersed), relative to those in the Separated condition (*answers* and *ideas* on separate pages), reported higher frequency of switching mindsets and lower willingness to engage and contribute content.

Besides due to weaker social recognition incentives, the negative effects on contribution are also consistent with the hypothesis that the integration of social content into the feed with expert knowledge content raised concerns among content contributors about diluting their professional image. In additional analyses of the field data, we find that integration caused a larger reduction in the posting of *ideas* that were unrelated (versus closely related) to *answers*; we reason that the content of closely related *ideas* is fairly more consistent with a professional expert image, while the content of unrelated *ideas* may not be. Also, we find that integration led to steeper decreases in the contribution of *ideas* among users who had more certified *answers* on the platform and among users who had given at least one paid talk on the platform—in other words, users who were highly invested in their professional image.

Our study contributes to the growing literature on the antecedents of online UGC by exploring how an important system design feature—the presentation format of heterogeneous types of UGC—affects how users contribute and engage with content. Our results suggest that the juxtaposition of content that requires different mindsets may decrease information processing fluency, thereby enriching the literature on consequences of the information presentation format on users' information processing and behaviors (e.g., Ives et al. 1980, Kim and Dennis 2019). We also contribute to the literature on how social image concerns affect online UGC generation (e.g., Wasko and Faraj 2005, Ma and Agarwal 2007, Goes et al. 2016, Qiu and Kumar 2017, Chen et al. 2018, Pu et al. 2020). We show that social image concerns seem to be heightened by merging disparate content into one channel, advancing our understanding of the image dilution effect in the decision science and marketing literature (Shanteau 1975, Nisbett et al. 1981, Pullig et al. 2015).

Practically, our results have important, concrete implications for all platforms that host (or are considering hosting) multiple types of UGC. Although additional content curation tools can enhance user engagement and stickiness under certain circumstances, our results suggest that community owners should maintain different information feeds for different types of content, at least if the content is heterogeneous. Otherwise, integration could adversely affect both user engagement and contribution behaviors, potentially threatening the longevity of the community as a whole.

## 2. Related Literature

Our research is closely related to the literature on the antecedents of user contributions and commitment to online platforms, which typically rely on people to contribute content voluntarily. Previous studies have found that people may be motivated to contribute because of knowledge self-efficacy, intrinsic enjoyment in helping others, social benefits and concerns, and monetary incentives (e.g., Wasko and Faraj 2005; Zhang and Zhu 2011; Goes et al. 2014, 2016; Chen et al. 2018; Pu et al. 2020; Wang et al. 2022).

Motivations related to social image are of particular interest in the literature on user contributions. Users may be motivated to contribute more and higher-quality knowledge to achieve a better reputation or status (Wasko and Faraj 2005, Goes et al. 2016). Online product reviewers may be strategic in their reviewing behaviors to attract the audience's attention (Shen et al. 2015). Important considerations and motivations include the group size of collaborators and audience members (Zhang and Zhu 2011, Toubia and Stephen 2013, Goes et al. 2014, Qiu and Kumar 2017, Baek and Shore 2020) and social tie density (Shriver et al. 2013). The addition of social networking functions to a platform can motivate users to contribute more content due to increased social presence (Huang et al. 2017).

Similar to in the offline world, people manage their virtual self-presentations to convey their desired image to their audience (Schau and Gilly 2003, Marwick and Boyd 2011, Belk 2013, Bullingham and Vasconcelos 2013, Marder et al. 2016). While the offline setting enables people to tailor their self-presentation to different audiences, separated by time and physical space (Goffman 1978), the virtual self is subject to simultaneous surveillance from diverse audience groups, causing a so-called *multiple audience problem* (Fleming et al. 1990, Marwick and Boyd 2011, Lee et al. 2015, Marder et al. 2016, Gil-Lopez et al. 2018). Previous research has shown that merging disparate audiences can affect users' online presentation behaviors. Specifically, consumers express more balanced opinions online if heterogeneous ratings are observed by multiple groups of audiences (Schlosser 2005, Moe and Trusov 2011). The merging of disparate audiences led Facebook users to adjust their presentation strategies and language styles (Gil-Lopez et al. 2018), perhaps displaying a *lowest common denominator effect*: people present themselves according to the standards of the strictest audience (Hogan 2010). For example, students might delete inappropriate images on Facebook if their teachers "friended" them.

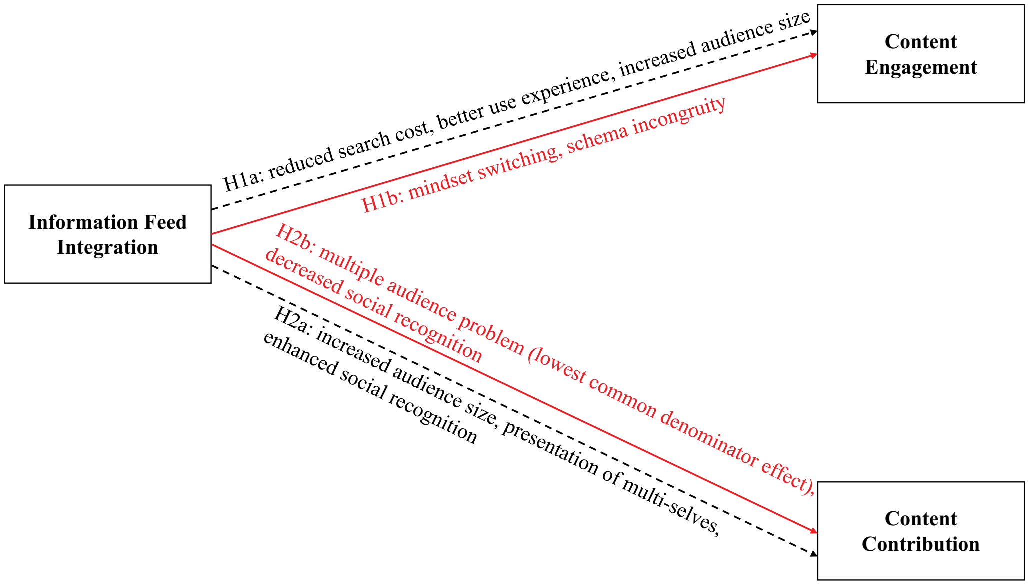
Our study also is broadly related to the literature on how the information presentation format influences users' judgments and decisions, for example, through information acquisition costs and processing strategies (Bettman and Kakkar 1977). The presentation format of financial statements influences the judgments and decisions of nonprofessional users (Maines and McDaniel 2000). The format of news on social media (source-primacy versus headline-primacy) affects the extent to which users believe the news as well as their subsequent engagement behaviors (Kim and Dennis 2019). Suboptimal display formats may induce cognitive strain (Bettman et al. 1991).

Although the decision of whether to mix or separate heterogeneous types of UGC represents a general design feature, the extant literature lacks an understanding of the implications of the decision for content consumers and content contributors. We attempt to fill this gap in knowledge about the consequences of information feed integration for both the information processing experiences of content consumers and the online self-presentation management of content contributors.

## 3. Hypothesis Development

We organize the hypothesis development by the outcome variables of our focal interests: content engagement and content contribution. Figure 2 summarizes the hypotheses we propose, which shall be elaborated next, and the theoretical lens underlying the hypotheses.

**Figure 2.** (Color online) Summary of Hypotheses and Theoretical Lens




## 3.1. Information Processing and Content Engagement

We start our hypothesis development about the effects of the information feed integration on users' information processing and content engagement. Note that in our natural experiment, "integration" involved deleting the separate information feed for *ideas* and merging all *ideas* into the main information feed (previously for *answers* and other expert knowledge content, e.g., *articles*, only).[4]

We will begin by reviewing the evidence in favor of positive effects of integration on user engagement with *ideas* and *answers*. We measure "engagement" with *ideas* as the number of likes and comments received by *ideas* and "engagement" with *answers* as the number of vote-ups, comments, and thanks received by *answers*. Before integration, users who wished to read *ideas* or *answers* had to navigate to separate channels. After integration, both types of content were shown to everyone who browsed the main channel, so the audience size increased for either content type, as did the ease of access—both of which could have boosted engagement with *ideas* and *answers* (Palmer 2002, Agnew and Szykman 2005). Moreover, the increased diversity and richness of information within the same channel could have improved users' information-acquisition experiences (Wu et al. 2019, Esfandiari et al. 2021), and more-satisfied users tend to engage more with both content types (Ma and Agarwal 2007). In sum, we propose:

**Hypothesis 1a.** *Ideas and answers received higher levels of user engagement after (versus before) information feed integration.*

Now, we will present the evidence in support of the opposite hypothesis that integration reduced user engagement with *ideas* and *answers*.

*Ideas* and *answers* are very different types of content. As shown in the Introduction, *answers* (posted in the Q&A forum) tend to be structured, in-depth and address specific questions, whereas *ideas* (originally posted in a separate channel) are free-form, short and shared for social interaction purposes. Given the vast differences between *ideas* and *answers*, we draw on mindset theory to theorize how integration might affect users' information processing of the two content types.

According to mindset theory (French II 2016), users employ different mindsets (or mental states) to accomplish different tasks (i.e., processing different types of content here). *Answers* and *ideas* likely require different mindsets to process, because the two types of content typically include different topics, have different structures (a structured response to a question versus unstructured casual information), and require different types of thinking modes, that is, *answers* typically require serious, deep thinking, while *ideas* require fast, intuitive thinking (Kahneman 2011). Of course, two *ideas* might be different enough to require different mindsets. On average, though, mindsets are likely to be more heterogeneous *between* the two types of content than *within* either type of content.

Moreover, building on the core concept of mindset theory, previous research has shown that when users process information, frequently switching between mindsets can deplete cognitive resources and impose extra psychic costs on users (Hamilton et al. 2011, Yan et al. 2018). When one type of content is interspersed with another quite different type of content in the information feed, it could break the continuity of the user's cognitive focus on processing one content type (Kahneman 1973; Speier et al. 1999, 2003), decreasing the user's speed and depth of information processing of the content (Engel et al. 1968). Consequently, users may have to conduct more selective information searches to obtain the desired content while coping with the higher cognitive load (Ask and Granhag 2005). Here, the integration caused the juxtaposition of content that requires different mindsets for processing, which likely forced users to switch more frequently between mindsets, thereby increasing the cognitive strain of information processing and inducing more selective information searches, which would be expected to reduce users' willingness to consume and engage with both types of content.

The reasoning from mindset theory presented above can also be supported by a similar theory in cognitive psychology—schema theory. According to schema theory, people build different mental structures (i.e., schemas) to organize, pay attention to, and process different categories of information (Bartlett 1932, Axelrod 1973, Alba and Hasher 1983). When users are confronted with content that requires inconsistent schema, their information processing can be negatively affected due to the *schema incongruity effect* (Meyers-Levy and Tybout 1989, Wheeler et al. 2005, Aggarwal and McGill 2007, Schulze et al. 2014).

In summary, we hence propose Hypothesis 1b to compete with Hypothesis 1a:

**Hypothesis 1b.** *Ideas and answers received lower levels of user engagement after (versus before) information feed integration.*

## 3.2. Social Recognition, Self-Presentation and Content Contribution

Integration may have boosted the contribution volume of *ideas* and *answers*, for three reasons. First, the incentive to contribute is known to be affected by the audience size (Goes et al. 2014, Qiu and Kumar 2017), which potentially increased for either content type after integration as both content types were shown to everyone who browsed the main channel, including those who previously had been unaware of either content type (especially the *ideas* content, which was introduced several years after the launch of the platform). Second, if integration boosted engagement with *ideas* and *answers* (as explained in Hypothesis 1a), then it also likely increased the incentive to contribute *ideas* and *answers* by enhancing social recognition benefits (von Krogh et al. 2012, Chen et al. 2018, Kuang et al. 2019). Third, online users have natural incentives to manage their impressions (Schau and Gilly 2003, Belk 2013). Users may wish to present multiple selves or personas on the same digital medium (Schau and Gilly 2003), so merging *ideas* into the same feed as *answers* might have enabled contributors to build a more complete persona, with both professional and personal components, hence enhancing their contribution incentives of both content types. Formally, we have the following Hypothesis 2a:

**Hypothesis 2a.** *Users contributed more ideas and answers after (versus before) information feed integration.*

When developing Hypothesis 2a, we introduced some research on how online users strategically manage their virtual self-presentations (Schau and Gilly 2003, Belk 2013). Other findings in the same body of literature support the opposite prediction: integration may *decrease* the incentive to contribute *ideas*, the social-interaction-oriented content.

In many online settings, unlike most offline settings (Goffman 1978), users face a *multiple audience problem* in which the presented self is subject to simultaneous surveillance by multiple audience groups (Fleming et al. 1990, Schlosser 2005, Marwick and Boyd 2011, Lee et al. 2015, Marder et al. 2016, Gil-Lopez et al. 2018). Along this logic, previous research further showed that when disparate audiences merge, users tend to choose presentation strategies that adhere to the standards of the strictest audience—the so-called *lowest common denominator effect* (Hogan 2010, Marder et al. 2016).

In our setting, information feed integration may have exacerbated the multiple audience problem by eliminating the separate space in which the *ideas* content was generated for and consumed by a smaller group of users than in the main information feed. After integration, *ideas* were seen by all—including those who were interested in only the expert knowledge content such as *answers*. The merging of audiences may have raised concerns that posting *ideas* would dilute the user's professional image (Shanteau 1975, Nisbett et al. 1981). Therefore, to meet the strictest standards for professional knowledge experts, users may have decreased their contributions of *ideas*. Indeed, studies have shown that users may strategically choose to post types of content (in this case, to reduce posting of *ideas*) that are congruent with their preferred social image (Wasko and Faraj 2005, Ma and Agarwal 2007, Goes et al. 2016, Qiu and Kumar 2017, Chen et al. 2018, Pu et al. 2020).

Finally, if users engage less with *ideas* and *answers* after integration due to the negative effects of frequent mindset switching on information processing as we explained in Hypothesis 1b, then users also may be demotivated to contribute *ideas* and *answers* due to decreased social recognition benefits (von Krogh et al. 2012, Chen et al. 2018, Kuang et al. 2019).

Overall, the reasoning presented above leads us to Hypothesis 2b, a competing hypothesis against Hypothesis 2a:

**Hypothesis 2b.** *Users contributed fewer ideas and answers after (versus before) information feed integration.*

# 4. Context and Data

## 4.1. Context

Launched on January 26, 2011, Zhihu is now one of the most popular Q&A communities in China. At the time of writing, it has more than 44 million questions and 240 million answers.[5] On August 24, 2017, Zhihu launched a new feature called *Ideas*, which allows users to share information and content that do not conform to the Q&A format.

Zhihu introduced the *Ideas* feature to increase user engagement and contributions. While *answer* writing is fairly professional and occurs in response to questions, users can write an *idea* without a pre-existing question, so the threshold of content generation is lower. Due to its free-form nature, users often share casual thoughts and personal life updates in *Ideas*, so it is similar to common social-networking services offered by Facebook and Twitter. Users can like and comment on the *ideas* posted by other users and the posting users can also reply to the comments. Users can see the *ideas* posted by their followees in their news feeds. An example of an *idea* shared by a user on Zhihu is shown in the left panel of Figure 3.

For almost two years after the introduction of *Ideas*, Zhihu presented all user-created *ideas* in a separate channel as indicated by the *Idea* tab in the left panel of Figure 3. Meanwhile, all the Q&A content and other expert knowledge content (e.g., *articles*, long expert knowledge content shared without targeting at specific questions) were accessible via the *Homepage* tab. On June 11, 2019, Zhihu integrated the *Idea* feed into the *Homepage* feed as shown in the right panel of Figure 3. Specifically, *ideas* were added to the *Homepage* subchannel *Following*, which presents all the content posted by the user's followees in a single feed, chronologically. The *Recommendation* and *Top List* subchannels within the *Homepage* tab were not affected by the change. (*Recommendation* presents *answers* recommended by the platform, while *Top List* presents the day's most popular *answers*.) Therefore, after integration, the most prominent change was that users now read both expert knowledge content and social content in the *Following* channel.

This nonselective and exogenous integration of heterogeneous content types provides a natural experiment with which we can examine how mixing social content with expert knowledge content affects user engagement and contribution behaviors. Figure 4 shows the relevant timeline and our sample window.

## 4.2. Data

We collected data from February 19 to October 1, 2019, so the sample period includes 16 weeks both before and after the focal event date (June 11, 2019). The data collection was started and finished in December 2020. We choose this time window because no other major events or updates happened during it, and it excludes major holidays, when users' contribution patterns tend to be dramatically different (e.g., Chinese Spring Festival, National Day Week; Kuang et al. 2019). Within our relatively narrow window, we can cleanly identify the effects of information feed integration.

To obtain a representative sample, we collected historical data from active users with a snowball sampling strategy (e.g., Goodman 1961, Bonaccorsi et al. 2006, Song et al. 2019, Wang et al. 2022). Specifically, we randomly selected five (seed) users who were active in posting both *answers* and *ideas* (i.e., users who frequently posted *answers* and *ideas* at the time of collecting the data). We obtained those users' followees, and then we obtained their followees, and so on until we collected 20,000 users. Finally, we randomly selected 5,000 out of the 20,000 users. Out of the randomly selected 5,000 users, there are 4,811 unique users. We collected the 4,811 users' publicly available information on Zhihu.

Given our research objects concerning the consequences of mixing two content types in one single information feed, we focus on the group of users who contributed both *ideas* and *answers* before the start of our observation window (i.e., February 19, 2019). If a user contributed both content types, this meant the user cared about both content types and was most likely to be sensitive to the design change (i.e., the information feed integration). We also excluded organizational accounts since their behaviors may systematically differ from individual users. The final sample used for analyses contained 1,892 users. In Online Appendix B, we present evidence that our sample is representative of the user population on Zhihu.

We collected information about various user activities including posting *ideas*, posting *answers*, liking *answers*, asking questions, and following questions. For each *idea* and *answer* posted by a user in our sample, we collected the posting time, number of vote-ups, number of likes, number of comments, and the text of the post. For the other types of activities, we obtained the activity's timestamp.

We converted the raw data into a user-week panel data set by aggregating the volumes of each activity (e.g., posting *ideas*, posting *answers*, liking *answers*) for each user each week. The panel data set has $1{,}892 \times 32(\text{weeks}) = 60{,}544$ observations for each variable. Moreover, to assess the impacts on user engagement, we calculated the average number of likes, vote-ups, and comments on all the *ideas* and *answers* posted by each user each week. Note that these measures are undefined for weeks in which the user did not post any content, so there are fewer than 60,544 observations for each of these variable. Table 2 presents the summary statistics and definitions of the variables.

**Figure 3.** (Color online) Zhihu *Ideas* Integration: Before (Left) vs. After (Right)

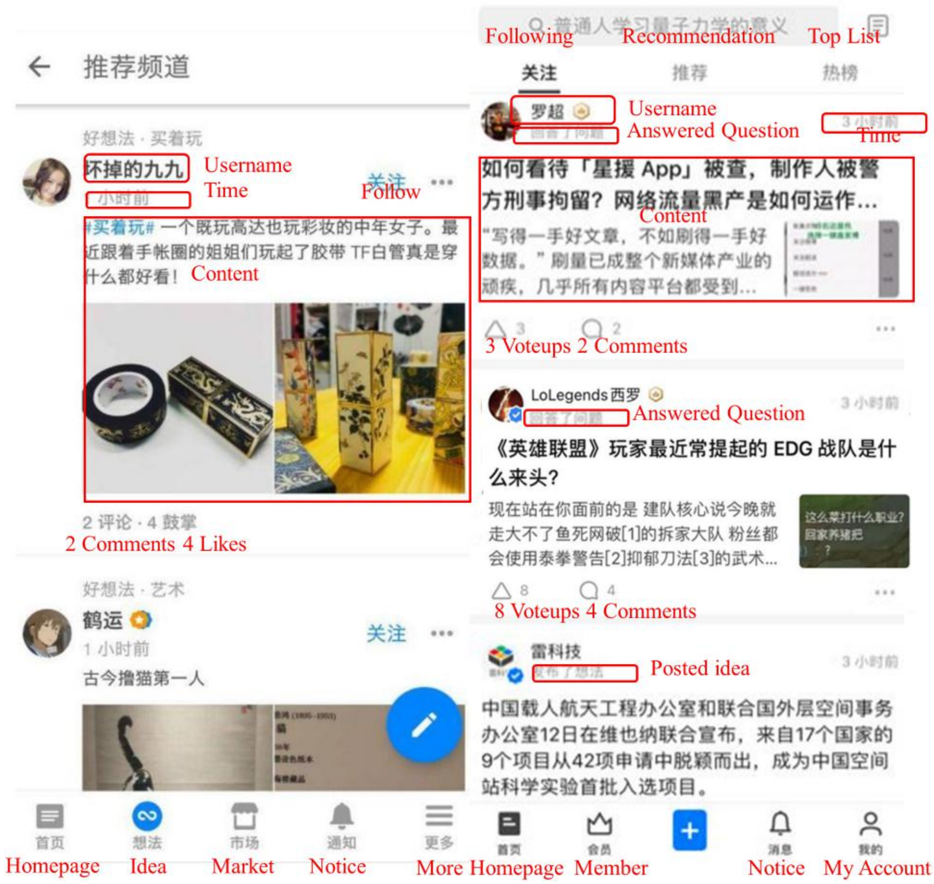
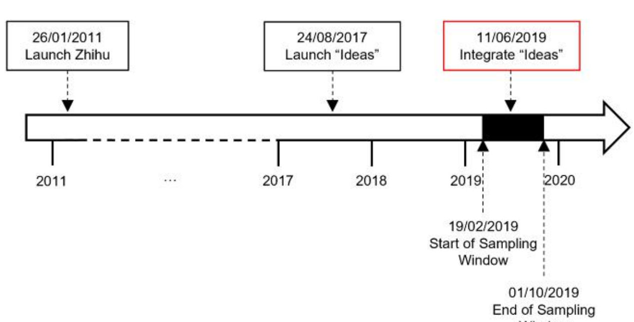



Table 3 compares the textual characteristics of *ideas* and *answers*. *Ideas* on average contained significantly fewer characters, figures, and links than *answers*, supporting that *answers* are longer and more in-depth than *ideas*.

In Table 4, we compare the outcomes of interest for *ideas* and for *answers* before and after feed integration. For *ideas*, the model-free evidence shows significant decreases in user engagement with and the contribution of *ideas* after integration. Figure 5 plots user engagement and contribution for *ideas*, specifically. The decreases in user engagement with and user contribution of *ideas* occurred immediately after integration. We also found directional to significant decreases in user engagement with and the contribution of *answers* after feed integration as shown in Table 4. The plots in Figure A3, Online Appendix A also show that user engagement with and user contribution of *answers* decreased after integration, but with noticeably smaller magnitudes.

Next, we test our hypotheses in a regression framework so that we can control for confounds and offer stronger causal inferences about the consequences of information feed integration.

**Figure 4.** (Color online) Timeline: Launch of *Ideas*, Integration, and Sampling




## 5. Main Analysis

### 5.1. Empirical Strategy

We use the regression in Equation (1) as our baseline model of the effect of information feed integration on user engagement and contribution behaviors:

$$y_{it} = \beta_0 + \beta_1 Integration_t + \gamma_1 Age_{it} + \gamma_2 Age_{it}^2 + \theta \mathbf{X}_{it} + u_i + \epsilon_{it}, \tag{1}$$

where $y_{it}$ is the outcome of interest (one of the variables listed in Table 4) for user $i$ in week $t$. $Integration_t$ is a dummy that equals 1 (0) if week $t$ is after (before)

**Table 2.** Descriptive Statistics

| Variables | Description | N | Mean | SD | Min | Max |
|---|---|---|---|---|---|---|
| $N\_ideas_{it}$ | The number of *ideas* contributed by user $i$ in week $t$ | 60,544 | 1.177 | 3.825 | 0 | 78 |
| $Avg\_comm\_idea_{it}$ | The average number of comments per *idea* for all *ideas* contributed by user $i$ in week $t$ | 15,368 | 5.49 | 14.713 | 0 | 1,213 |
| $Avg\_like\_idea_{it}$ | The average number of likes per *idea* for all *ideas* contributed by user $i$ in week $t$ | 15,368 | 23.015 | 45.556 | 0 | 1,884 |
| $Avg\_char\_idea_{it}$ | The average number of characters per *idea* for all *ideas* contributed by user $i$ in week $t$ | 15,368 | 65.374 | 85.177 | 0 | 1,697 |
| $Avg\_link\_idea_{it}$ | The average number of links per *idea* for all *ideas* contributed by user $i$ in week $t$ | 15,368 | 0.289 | 0.493 | 0 | 10 |
| $N\_ans_{it}$ | The number of *answers* contributed by user $i$ in week $t$ | 60,544 | 1.253 | 4.464 | 0 | 156 |
| $Avg\_vote\_ans_{it}$ | The average number of vote-ups per answer for all *answers* contributed by user $i$ in week $t$ | 18,911 | 530.645 | 3,390.094 | 0 | 245,914 |
| $Avg\_comm\_ans_{it}$ | The average number of comments per answer for all *answers* contributed by user $i$ in week $t$ | 18,911 | 46.086 | 232.448 | 0 | 22,252 |
| $Avg\_thank\_ans_{it}$ | The average number of thanks per answer for all *answers* contributed by user $i$ in week $t$ | 18,911 | 110.408 | 1,406.448 | 0 | 116,675 |
| $Avg\_char\_ans_{it}$ | The average number of characters per answer for all *answers* contributed by user $i$ in week $t$ | 18,911 | 730.380 | 1,115.493 | 0 | 23,841 |
| $Avg\_fig\_ans_{it}$ | The average number of figures per answer for all *answers* contributed by user $i$ in week $t$ | 18,911 | 2.669 | 6.913 | 0 | 179 |
| $Avg\_link\_ans_{it}$ | The average number of links per answer for all *answers* contributed by user $i$ in week $t$ | 18,911 | 0.593 | 2.468 | 0 | 115 |
| $N\_ans\_liked_{it}$ | The number of *answers* liked by user $i$ in week $t$ | 60,544 | 7.991 | 29.168 | 0 | 923 |
| $N\_ques\_asked_{it}$ | The number of questions asked by user $i$ in week $t$ | 60,544 | 0.044 | 0.427 | 0 | 42 |
| $N\_ques\_followed_{it}$ | The number of questions followed by user $i$ in week $t$ | 60,544 | 1.372 | 7.001 | 0 | 239 |
| $Age_{it}$ | Number of 100 weeks since user $i$'s first action on Zhihu | 60,544 | 2.303 | 0.992 | 0.07 | 4.57 |
| $Age_{it}^2$ | Square of $Age_{it}$ | 60,544 | 6.287 | 4.821 | 0.005 | 20.885 |

*Note.* There are fewer observations for variables regarding content engagements, because those variables are undefined for weeks in which the user posted no content.

information feed integration. $u_i$ captures user fixed effects, which control for time-invariant user heterogeneity that might confound the estimation of the effect of integration on the outcome of interest.

Since information feed integration affected all users, we lack a control group of users with which we could infer the counterfactual behaviors of users in the treated group. In other words, the most salient confounder of the estimated effect of $Integration_t$ in the baseline model (1) is that user behaviors may vary over time in ways that do not relate to the shock. In the next paragraphs, we explain how we strengthened causality by including several time-varying covariates in the baseline model and by using a regression discontinuity in time (RDiT) framework to control for potentially different nonlinear time trends before and after the shock. Also, in Section 5.4, we conduct more robustness checks, including a random trend model that controls for individual-specific time trends and a falsification test using a placebo event date to ensure that our results are not driven by spurious correlations.

**Table 3.** Comparison Between *Ideas* and *Answers*

| Content | *Ideas* | | *Answers* | | T-test |
|---|---|---|---|---|---|
| Characteristics | Mean | SD | Mean | SD | *t*-stats |
| No. of characters | 65.374 | 85.177 | 730.380 | 1,115.493 | −64.79*** |
| No. of figures | 1.088 | 1.557 | 2.669 | 6.913 | −26.25*** |
| No. of links | 0.289 | 0.493 | 0.593 | 2.468 | −14.31*** |

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

To ensure that our estimated effects of integration are not driven by natural trends in user behavior over time, we include user age, defined as the number of weeks since the user's first action on the platform, and the square of user age to allow for nonlinear effects (Zhang and Zhu 2011). We also include some time-varying covariates, denoted $\mathbf{X}_{it}$, that may correlate with contribution and engagement behaviors in the regression model. Specifically, we add the number of *answers* liked, number of questions asked, and number of questions followed. $\theta$ represents the vector of coefficients for the control variables. $\epsilon_{it}$ is the random error term. We use the logarithm of all continuous variables (adding 1 to the original value when necessary) since the dependent variables are quite skewed.

To estimate the effects in the RDiT framework (Davis 2008, Auffhammer and Kellogg 2011, Anderson 2014, Hausman and Rapson 2018), we use the duration as the running variable and the treatment date (i.e., June 11, 2019) as the discontinuity threshold. The RDiT approach is used in many disciplines to estimate the causal impacts of policy changes (Hausman and Rapson 2018).

**Table 4.** Summary Statistics for Engagement and Contribution Before and After Integration

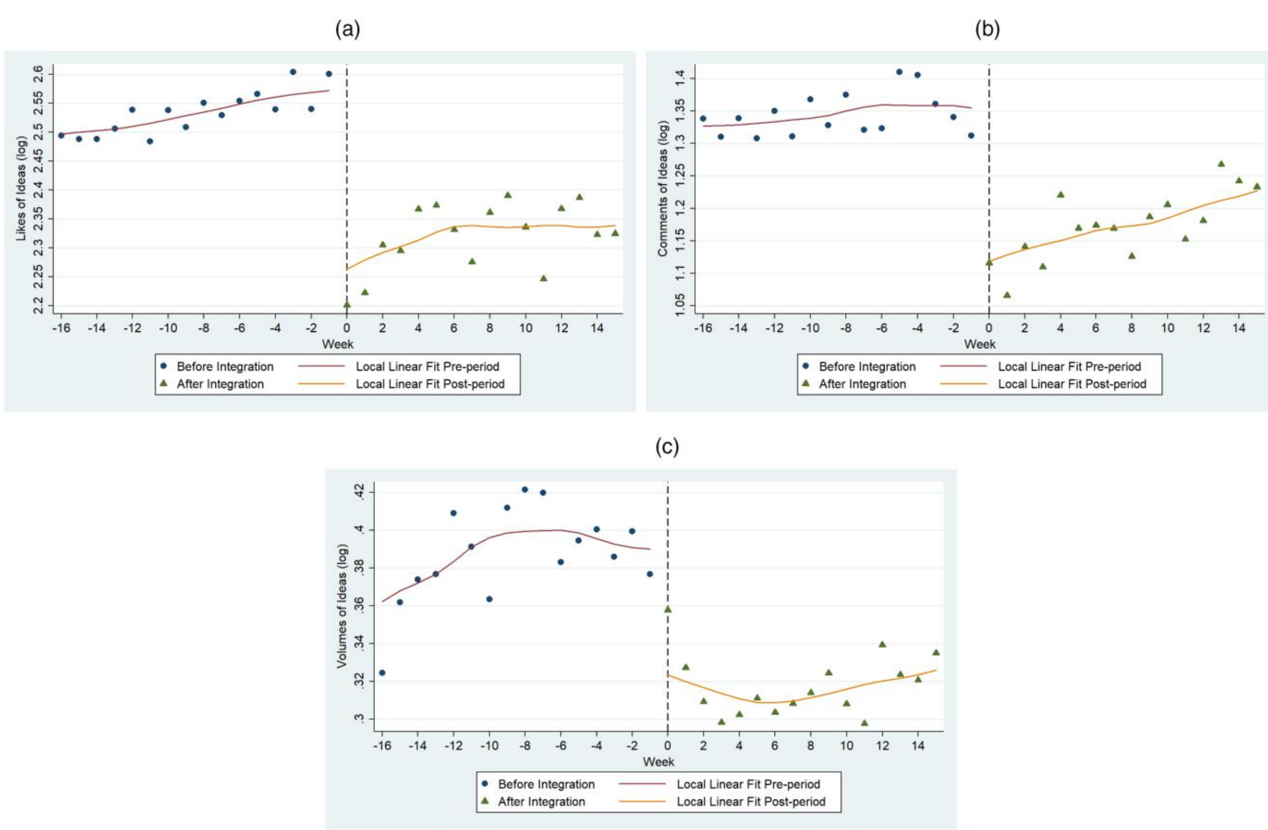
| | Pre-integration | | Post-integration | | Difference | Paired $t$-test |
|---|---|---|---|---|---|---|
| Variables | Mean (1) | Standard error (2) | Mean (3) | Standard error (4) | Value (5) | $t$-stats (6) |
| Volumes of *Ideas* (log) | 0.3871 | 0.0043 | 0.3176 | 0.0039 | −0.0695 | −20.7658*** |
| Comments on *Ideas* (log) | 1.4574 | 0.0147 | 1.2871 | 0.0145 | −0.1702 | −14.3396*** |
| Likes on *Ideas* (log) | 2.6611 | 0.0172 | 2.4817 | 0.0173 | −0.1794 | −18.4549*** |
| Volumes of *Answers* (log) | 0.4004 | 0.0040 | 0.3995 | 0.0041 | −0.0008 | −0.2301 |
| Vote-ups on *Answers* (log) | 3.8374 | 0.0286 | 3.8029 | 0.0285 | −0.0345 | −1.3386 |
| Comments on *Answers* (log) | 2.3768 | 0.0216 | 2.3332 | 0.0216 | −0.0436 | −1.9963** |
| Thanks on *Answers* (log) | 2.2103 | 0.0234 | 2.1383 | 0.0233 | −0.0721 | −3.3373*** |

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

Traditional cross-sectional regression discontinuity (RD) designs exploit a policy that applies only to units that are above or below a certain threshold; the treatment effect is estimated by comparing the conditional means of the cross-sectional units just above and below the threshold (Hausman and Rapson 2018). The traditional RD design is not applicable in our setting since information feed integration affected all users. Meanwhile, in RDiT, the identifying variation is discontinuity in time within a short time window, which is evident in Figure 5. The approach is similar to that in an interrupted time series or a simple pre/post comparison (Hausman and Rapson 2018), but RDiT has the advantage of flexibly controlling for nonlinear time trends before and after the treatment shock, thus strengthening the causal interpretation of the estimations (Davis 2008). We also conduct a robustness check regarding this assumption using the "augmented local linear" approach (Hausman and Rapson 2018) in Section 5.4.

Consistent with prior studies (e.g., Goes et al. 2016, Lee et al. 2018), we specify the RDiT model as a

**Figure 5.** (Color online) Engagement with and Contribution of *Ideas* Before vs. After Integration




*Notes.* (a) Likes on *Ideas*. (b) Comments on *Ideas*. (c) Contribution volumes of *Ideas*.

parametric polynomial model with user fixed effects:

$$y_{it} = \beta_0 + \beta_1 Integration_t + \sum_{p=0}^{P} \beta_{2,p} Duration_t^p + \sum_{p=0}^{P} \beta_{3,p} Integration_t \times Duration_t^p + \gamma_1 Age_{it} + \gamma_2 Age_{it}^2 + \theta \mathbf{X}_{it} + u_i + \epsilon_{it}, \quad (2)$$

where $Duration_t$, the running variable, is the number of weeks after or before information feed integration (i.e., the absolute value of the difference in weeks between week $t$ and the shock week). The interaction term $Integrationt \times Duration^p$ allows the regression function to differ on either side of the cutoff point (Lee and Lemieux 2010, Goes et al. 2016). We also vary the maximum polynomial orders (i.e., the value of $P$) from 1 to 3 to assess the robustness of the RDiT estimation results. All other notations in Equation (2) are the same as in Equation (1). Note also that Equation (1) is a special case of Equation (2) wherein $p = 0$.

### 5.2. Effects on Engagement

In Table 5, we present the estimated effects of information feed integration on engagement with *ideas*. In Column (1), the estimated effect of integration on *Avg_comm_idea* is significantly negative (–0.2693, $p < 0.01$); integration led to an average decrease of 24% in the number of comments on *ideas*.[6] In Column (2), the estimated effect of integration on *Avg_like_idea* is significantly negative (–0.3278, $p < 0.01$); integration led to an average decrease of 28% in the number of likes on *ideas*.

**Table 5.** Effects of Information Feed Integration on User Engagement with *Ideas*

| Variables | Comments (1) | Likes (2) | Length (3) | Links (4) |
|---|---|---|---|---|
| *Integration* | −0.2693*** | −0.3278*** | −0.0224 | −0.0124 |
| | (0.022) | (0.018) | (0.030) | (0.008) |
| *Age* | 0.8875*** | 1.2023*** | −0.0850 | −0.0231 |
| | (0.238) | (0.235) | (0.298) | (0.095) |
| $Age^2$ | −0.0486 | −0.0631 | −0.0612 | 0.0117 |
| | (0.043) | (0.043) | (0.052) | (0.016) |
| *N_ans_liked* | 0.0199** | −0.0056 | 0.0470*** | 0.0181*** |
| | (0.010) | (0.008) | (0.015) | (0.004) |
| *N_ques_asked* | −0.0305 | −0.0061 | 0.0387 | 0.0143 |
| | (0.028) | (0.025) | (0.045) | (0.011) |
| *N_ques_followed* | 0.0069 | −0.0125 | 0.0385** | 0.0089** |
| | (0.012) | (0.010) | (0.018) | (0.005) |
| *Constant* | −0.3897 | 0.2229 | 4.1126*** | 0.1543 |
| | (0.347) | (0.322) | (0.448) | (0.139) |
| Observations | 15,368 | 15,368 | 15,368 | 15,368 |
| User fixed effects | Yes | Yes | Yes | Yes |
| Adjusted $R^2$ | 0.659 | 0.851 | 0.418 | 0.395 |
| Number of users | 1,257 | 1,257 | 1,257 | 1,257 |

*Notes.* Robust standard errors are in parentheses. For each user, the observations for weeks when the user posted no *ideas* are dropped because engagement metrics are undefined in these weeks, which leads to the number of users used in this analysis less than 1,892.
***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

In Columns (3) and (4) of Table 5, we examine how integration affected the length (*Avg_char_idea*) and number of links (*Avg_link_idea*) in *ideas* as proxies for the quality of the content (Yaari et al. 2011, Burtch et al. 2017, Wang et al. 2022). Unlike in Columns (1) and (2), the estimated effects of *Integration* are statistically insignificant. Moreover, using text analysis, we find integration had no impact on *ideas*'s text readability and cognitive processing language (see Table D1 in Online Appendix D). We conclude that the decrease in user engagement with *ideas* is not likely driven by a decrease in the quality of the content.

In Table 6, we present the estimated effects of information feed integration on engagement with *answers*. In Columns (1) through (3), the estimated effects are significant and negative: integration led to average decreases of 10% in the number of vote-ups (*Avg_voteup_ans*; Column (1)), 7% in the number of comments (*Avg_comm_ans*; Column (2)), and 9% in the number of thanks (*Avg_thank_ans*; Column (3)) received by the *answers* in our sample.

In Columns (4) through (6) of Table 6, we examine how integration affected the objective quality of the *answers*. As in Table 5, we find no significant effects of integration on the number of characters, figures, or links in *answers*. Moreover, based on text analysis, we find integration also had no impact on *answers*'s text readability and cognitive processing language (see Table D2 in Online Appendix D). So, the decrease in user engagement with *answers* is not likely due to deteriorated content quality.

The results presented in this subsection support our Hypothesis 1b and enable us to reject Hypothesis 1a.

We assess the robustness of the effects using the RDiT framework in Equation (2). The results, presented in Online Appendix C, are largely consistent with the results presented in Tables 5 and 6, with one noticeable exception: When we included higher orders (2 or 3) of the polynomial terms of the running variable, $Duration_t$, the estimated effects of integration on engagement with *answers* became insignificant (though the signs remain negative). This is unsurprising given the relative magnitudes of the effects in Tables 5 and 6—the estimated coefficients in the first two columns of Table 5 (regarding *ideas*) are larger in magnitude than the coefficients in the first three columns of Table 6 (regarding *answers*).

Finally, we estimate the effects of integration on engagement with *ideas* and *answers* simultaneously using seemingly unrelated regression (SUR). Again, the estimated coefficients of *Integration* are larger in the *ideas* equation than in the *answers* equation (see Tables D3 and D4 in Online Appendix D).

**Table 6.** Effects of Information Feed Integration on User Engagement with *Answers*

| Variables | Vote-ups (1) | Comments (2) | Thanks (3) | Length (4) | Figures (5) | Links (6) |
|---|---|---|---|---|---|---|
| *Integration* | −0.1039** | −0.0727* | −0.0998*** | −0.0151 | 0.0254 | −0.0058 |
| | (0.046) | (0.038) | (0.038) | (0.031) | (0.018) | (0.012) |
| *Age* | 0.6944 | 0.2336 | 0.2901 | 0.0649 | 0.1876 | 0.1436 |
| | (0.449) | (0.368) | (0.371) | (0.311) | (0.192) | (0.118) |
| *Age*$^2$ | −0.0565 | −0.0146 | −0.0373 | −0.0515 | −0.0428 | −0.0323 |
| | (0.080) | (0.068) | (0.066) | (0.056) | (0.034) | (0.022) |
| *N_ans_liked* | 0.2004*** | 0.1527*** | 0.1393*** | 0.0857*** | 0.0163* | 0.0128** |
| | (0.021) | (0.018) | (0.017) | (0.014) | (0.009) | (0.005) |
| *N_ques_asked* | −0.0031 | −0.0387 | −0.0494 | −0.0164 | −0.0099 | 0.0070 |
| | (0.052) | (0.044) | (0.045) | (0.047) | (0.024) | (0.019) |
| *N_ques_followed* | 0.1144*** | 0.0860*** | 0.0986*** | 0.0883*** | 0.0324*** | 0.0212** |
| | (0.026) | (0.022) | (0.022) | (0.020) | (0.011) | (0.009) |
| *Constant* | 2.1656*** | 1.5738*** | 1.4642*** | 5.6724*** | 0.4419 | 0.0891 |
| | (0.646) | (0.524) | (0.540) | (0.444) | (0.269) | (0.169) |
| Observations | 18,911 | 18,911 | 18,911 | 18,911 | 18,911 | 18,911 |
| User fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted $R^2$ | 0.603 | 0.501 | 0.575 | 0.558 | 0.605 | 0.442 |
| Number of Users | 1,490 | 1,490 | 1,490 | 1,490 | 1,490 | 1,490 |

*Notes.* Robust standard errors in parentheses. For each user, the observations for weeks when the user posted no *answers* are dropped because engagement metrics are undefined in these weeks, which leads to the number of users used in this analysis less than 1,892.
***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

Together, the results of the RDiT and SUR analyses suggest that integration affected user engagement with *ideas* more than with *answers*. This might be because of the fact that the integration here involved deleting the separate information feed for *ideas* and merging all *ideas* into the main information feed, previously for *answers* and other expert knowledge content only. Hence, most users are more likely to ignore *ideas*, the less-known content type, in the integrated information feed.

## 5.3. Effects on Contribution

In this subsection, we present the estimated effects of integration on contribution. Table 7 presents the estimated effects on the number of *ideas* contributed (*N_ideas*); Column (1) presents the estimates from Equation (1), while Columns (2) through (4) present the estimates from Equation (2) with a maximum polynomial order of 1 to 3, respectively. All estimates show a significantly negative effect of integration on the volume of *ideas* contributed by the users in our sample. Specifically, the baseline estimate in Column (1) suggests a decrease of 8% in the weekly volume of *ideas*.

Table 8 presents the estimated effects on the number of *answers* contributed (*N_ans*); the structure is analogous to Table 7. The result in Column (1) indicates that integration led to a 2% decrease in the weekly volume of *answers* contributed by users in our sample. The result is also significant when the maximum polynomial order is 1 (Column (2)) but not when it is 2 or 3 (Columns (3) and (4)). This might be because the information feed integration discouraged the contribution of *answers* (versus *ideas*) less saliently. The SUR estimation also confirms that the coefficient of *Integration* is significantly larger in the *ideas* equation than in the *answers* equation (see Table D5 in Online Appendix D). A possible reason is that users' engagements with *ideas* (versus *answers*) decreased by a larger extent as documented in Section 5.2, leading to a greater reduction in social recognition benefits for *ideas* than for *answers*. Besides, the image dilution concern acts as an additional factor to discourage contribution of *ideas* (see the arguments for Hypothesis 2b in Section 3.2). Overall, these results presented in this subsection support Hypothesis 2b and reject Hypothesis 2a.

## 5.4. Robustness Checks

This subsection presents several robustness checks and falsification tests. First, since the RDiT approach relies on discontinuity within a short time window, we check the robustness of the RDiT approach by focusing on a shorter time window. Following Hausman and Rapson (2018), we use a two-step procedure called the "augmented local linear" approach. Specifically, in the first step, we regress the dependent variable on the fixed effects and control variables using the full sample of observations (i.e., 16 weeks before and after integration). In the second step, we estimate a local linear specification by regressing the residuals from the first step on the treatment indicator, *Integration*, within a shorter time window (8 weeks before and after integration). The results, in Table 9, are consistent with the main findings.

In the second robustness check, we extend the baseline model in Equation (1) to allow each user to have their own time trend. We estimate the following random

**Table 7.** Effects of Information Feed Integration on the Contribution of *Ideas*

| | DV: Number of *ideas* (*N_ideas*) | | | |
|---|---|---|---|---|
| Variables | Polynomial Order 0 (1) | Polynomial Order 1 (2) | Polynomial Order 2 (3) | Polynomial Order 3 (4) |
| *Integration* | −0.0841*** | −0.0852*** | −0.0327*** | −0.0452*** |
| | (0.009) | (0.009) | (0.010) | (0.012) |
| *Duration* | | −0.0011** | 0.0004 | −0.0113*** |
| | | (0.001) | (0.002) | (0.003) |
| *Duration*$^2$ | | | −0.0007*** | 0.0013* |
| | | | (0.000) | (0.001) |
| *Duration*$^3$ | | | | −0.0001*** |
| | | | | (0.000) |
| *Integration* × *Duration*$^2$ | | | 0.0012*** | 0.0009 |
| | | | (0.000) | (0.001) |
| *Integration* × *Duration*$^3$ | | | | 0.0000 |
| | | | | (0.000) |
| *Age* | 0.2256** | 0.2203** | −0.7452*** | −0.5332 |
| | (0.112) | (0.112) | (0.191) | (0.348) |
| *Age*$^2$ | −0.023 | −0.0219 | −0.0218 | −0.0218 |
| | (0.020) | (0.020) | (0.020) | (0.020) |
| *N_ans_liked* | 0.1048*** | 0.1048*** | 0.1044*** | 0.1044*** |
| | (0.007) | (0.007) | (0.007) | (0.007) |
| *N_ques_asked* | 0.1239*** | 0.1237*** | 0.1233*** | 0.1231*** |
| | (0.028) | (0.028) | (0.028) | (0.028) |
| *N_ques_followed* | 0.0901*** | 0.0902*** | 0.0901*** | 0.0901*** |
| | (0.009) | (0.009) | (0.009) | (0.009) |
| *Constant* | −0.1246 | −0.1102 | 2.0845*** | 1.6177** |
| | (0.152) | (0.151) | (0.373) | (0.769) |
| Observations | 60,544 | 60,544 | 60,544 | 60,544 |
| User fixed effects | Yes | Yes | Yes | Yes |
| Adjusted $R^2$ | 0.727 | 0.727 | 0.728 | 0.728 |
| Number of users | 1,892 | 1,892 | 1,892 | 1,892 |

*Notes.* The interaction term *Integration* × *Duration* is omitted because it is collinear with *Age*. Robust standard errors clustered by users are in parentheses.
***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

trend model (Wooldridge 2010, p. 375):

$$y_{it} = g_i t + \beta_0 + \beta_1 Integration_t + \gamma_1 Age_{it} + \gamma_2 Age_{it}^2 + \theta \mathbf{X}_{it} + u_i + \epsilon_{it}, \tag{3}$$

where $g_i t$ is the individual-specific time trend for user $i$. The estimation results, reported in Table 10, are consistent with the main results. This increases our confidence that our effects are driven by integration rather than by an unrelated trend in user behaviors.

To ensure that our results are not driven by spurious correlations, we conduct a falsification test. We focus on the 16-week period prior to integration, and we use 8 weeks before the actual date of integration as the placebo event, so the analysis does not include the real postintegration period (for a similar falsification test, see Bapna et al. 2018). We should not observe significant coefficients of *Integration* because there is no real integration.

Table 11 presents the estimation results for the contribution and engagement outcomes for *ideas* and *answers*. As expected, we find no significant effects of the placebo event, though we find a marginal, positive effect on the number of comments received by *answers* (Column (6)).

We conduct an analogous falsification test in the 16-week posttreatment period, using 8 weeks after the actual date of integration as the placebo event. The results are largely consistent (see Table D6 in Online Appendix D).

In Online Appendix D, we report additional robustness checks: the inclusion of *articles* (another type of expert knowledge content) alongside *answers* (Table D7), and SUR analyses to account for the correlations between engagements and contributions (Tables D8 and D9). All results are consistent with those reported in the main text.

## 6. Underlying Mechanisms

### 6.1. Content Engagement

In Section 5.2, we showed that integrating social posts (*ideas*) with expert knowledge content (*answers*) decreased user engagement with both types of content, though more so for *ideas* than for *answers*. We now test our theorization regarding the negative effects of integration on

**Table 8.** Effects of Information Feed Integration on the Contribution of *Answers*

| | DV: Number of *answers* (*N_ans*) | | | |
|---|---|---|---|---|
| Variables | Polynomial Order 0 (1) | Polynomial Order 1 (2) | Polynomial Order 2 (3) | Polynomial Order 3 (4) |
| *Integration* | −0.0197** | −0.0206** | 0.0121 | 0.0104 |
| | (0.008) | (0.008) | (0.010) | (0.013) |
| *Duration* | | −0.0009* | 0.0035** | 0.0042 |
| | | (0.001) | (0.002) | (0.004) |
| *Duration*$^2$ | | | −0.0006*** | −0.0006 |
| | | | (0.000) | (0.001) |
| *Duration*$^3$ | | | | −0.0000 |
| | | | | (0.000) |
| *Integration* × *Duration*$^2$ | | | 0.0007*** | 0.0004 |
| | | | (0.000) | (0.001) |
| *Integration* × *Duration*$^3$ | | | | 0.0000 |
| | | | | (0.000) |
| *Age* | 0.1313 | 0.1268 | −0.4312** | −0.3557 |
| | (0.098) | (0.098) | (0.193) | (0.395) |
| *Age*$^2$ | 0.0071 | 0.0081 | 0.0083 | 0.0083 |
| | (0.017) | (0.017) | (0.017) | (0.017) |
| *N_ans_liked* | 0.1676*** | 0.1676*** | 0.1674*** | 0.1674*** |
| | (0.009) | (0.009) | (0.009) | (0.009) |
| *N_ques_asked* | 0.1440*** | 0.1437*** | 0.1436*** | 0.1436*** |
| | (0.021) | (0.021) | (0.021) | (0.021) |
| *N_ques_followed* | 0.1563*** | 0.1563*** | 0.1562*** | 0.1562*** |
| | (0.010) | (0.010) | (0.010) | (0.010) |
| *Constant* | −0.1711 | −0.1590 | 1.0974*** | 0.9232 |
| | (0.140) | (0.141) | (0.398) | (0.884) |
| Observations | 60,544 | 60,544 | 60,544 | 60,544 |
| User fixed effects | Yes | Yes | Yes | Yes |
| Adjusted $R^2$ | 0.671 | 0.671 | 0.671 | 0.671 |
| Number of users | 1,892 | 1,892 | 1,892 | 1,892 |

*Notes.* The interaction term *Integration* × *Duration* is omitted because it is collinear with *Age*. Robust standard errors clustered by users are in parentheses.
***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

user engagement: Mixing different content types hurts users' information processing. Given the micro nature of the mechanism we propose (i.e., the mindset switching effect), we test it by conducting a user-level laboratory experiment. By doing so, we sought more direct evidence about the roles of information acquisition and processing. Moreover, by randomly allocating subjects into the treated and control conditions in the experiment, we could eliminate the endogeneity concerns that are inherent to the field setting, thereby establishing a clearer causal effect.

We used four *answers* and eight *ideas* from Zhihu as stimuli (see examples in Online Appendix E). We used a between-subjects design with two conditions: Separated and Mixed. In the Separated condition, subjects saw the four *answers* on one page followed by the eight

**Table 9.** RDiT Robustness Check: Augmented Local Linear Approach

| | *Ideas* | | | *Answers* | | | |
|---|---|---|---|---|---|---|---|
| Variables | Volume (1) | Comments (2) | Likes (3) | Volume (4) | Vote-ups (5) | Comments (6) | Thanks (7) |
| *Integration* | −0.0541*** | −0.1571*** | −0.1923*** | −0.0144*** | −0.0753*** | −0.0579*** | −0.0691*** |
| | (0.004) | (0.010) | (0.011) | (0.005) | (0.026) | (0.020) | (0.020) |
| *Constant* | 0.0308*** | 0.0635*** | 0.0833*** | 0.0116*** | 0.0351* | 0.0307** | 0.0249* |
| | (0.003) | (0.008) | (0.006) | (0.003) | (0.019) | (0.015) | (0.015) |
| Observations | 30,272 | 7,780 | 7,780 | 30,272 | 9,568 | 9,568 | 9,568 |
| $R^2$ | 0.006 | 0.018 | 0.043 | 0.000 | 0.001 | 0.001 | 0.001 |
| Number of users | 1,892 | 1,088 | 1,088 | 1,892 | 1,361 | 1,361 | 1,361 |

*Note.* Bootstrapped standard errors in parentheses.
***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

**Table 10.** Random Trend Model

| | *Ideas* | | | *Answers* | | | |
|---|---|---|---|---|---|---|---|
| Variables | Volumes (1) | Comments (2) | Likes (3) | Volumes (4) | Vote-ups (5) | Comments (6) | Thanks (7) |
| *Integration* | −0.0849*** | −0.2680*** | −0.3260*** | −0.0205** | −0.1010** | −0.0772** | −0.0944** |
| | (0.009) | (0.023) | (0.018) | (0.008) | (0.047) | (0.039) | (0.039) |
| *Age*$^2$ | −0.7170** | 1.5080* | 0.2740 | −0.7110** | −0.4490 | −0.1820 | 0.7700 |
| | (0.301) | (0.814) | (0.659) | (0.323) | (1.700) | (1.458) | (1.444) |
| *N_ans_liked* | 0.0933*** | 0.0025 | −0.0170** | 0.1510*** | 0.1810*** | 0.1370*** | 0.1180*** |
| | (0.006) | (0.010) | (0.008) | (0.008) | (0.021) | (0.018) | (0.018) |
| *N_ques_asked* | 0.1010*** | −0.0389 | −0.0097 | 0.1230*** | 0.0189 | −0.0249 | −0.0250 |
| | (0.024) | (0.030) | (0.024) | (0.020) | (0.056) | (0.048) | (0.048) |
| *N_ques_followed* | 0.0780*** | 0.0105 | −0.0064 | 0.1510*** | 0.1150*** | 0.0921*** | 0.0999*** |
| | (0.007) | (0.012) | (0.010) | (0.009) | (0.027) | (0.023) | (0.023) |
| *Constant* | 4.2110** | −7.0800 | 0.9490 | 4.0970** | 5.5920 | 2.9100 | −1.9530 |
| | (1.662) | (4.519) | (3.657) | (1.787) | (8.521) | (7.307) | (7.236) |
| Observations | 60,544 | 15,368 | 15,368 | 60,544 | 18,911 | 18,911 | 18,911 |
| User fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Individual-specific time trend | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.236 | 0.153 | 0.222 | 0.209 | 0.113 | 0.108 | 0.114 |
| Number of users | 1,892 | 1,257 | 1,257 | 1,892 | 1,490 | 1,490 | 1,490 |

*Notes.* *Age* is omitted because it is collinear with the individual-specific time trend. Robust standard errors clustered by user are in parentheses.
***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

*ideas* on the next page. The orders of the *answers* and *ideas* were randomized on their respective pages. In the Mixed condition, the *answers* and *ideas* were mixed together and presented in a randomized order across two pages (with six pieces of content per page).

We recruited 59 undergraduate students from a large public university in China. Each subject was contacted by the experimenter beforehand and asked to bring their laptops to a designated room to accomplish the experiment. Once subjects arrived at the experiment room, the experimenter briefly introduced the experiment and then sent them a Qualtrics survey link. Subjects completed the survey on their own computers. Once they clicked the link, subjects were randomly assigned to the two conditions, viewed the content as described above, and then answered questions about their information acquisition and processing experiences.

We borrowed two items from Lee et al. (2010) to measure processing fluency (*"The content I browsed was easy to digest and absorb"*, *"The content I browsed was*

**Table 11.** Falsification Test

| | *Ideas* | | | *Answers* | | | |
|---|---|---|---|---|---|---|---|
| Variables | Volumes (1) | Comments (2) | Likes (3) | Volumes (4) | Vote-ups (5) | Comments (6) | Thanks (7) |
| *Integration* | 0.0008 | 0.0216 | −0.0072 | 0.0144 | 0.0621 | 0.1077* | 0.0603 |
| | (0.009) | (0.029) | (0.021) | (0.010) | (0.064) | (0.055) | (0.054) |
| *Age* | 0.2391 | −0.2313 | 0.5306 | 0.0462 | 0.1494 | −0.5681 | −0.1448 |
| | (0.217) | (0.566) | (0.450) | (0.204) | (1.032) | (0.889) | (0.879) |
| *Age*$^2$ | −0.0026 | 0.0802 | 0.0672 | 0.0172 | −0.0810 | −0.0835 | −0.1407 |
| | (0.038) | (0.097) | (0.080) | (0.036) | (0.191) | (0.168) | (0.163) |
| *N_ans_liked* | 0.0986*** | −0.0005 | −0.0083 | 0.1558*** | 0.1787*** | 0.1427*** | 0.1204*** |
| | (0.008) | (0.014) | (0.011) | (0.010) | (0.033) | (0.027) | (0.027) |
| *N_ques_asked* | 0.1198*** | 0.0036 | 0.0104 | 0.1356*** | −0.0219 | −0.0411 | −0.0371 |
| | (0.029) | (0.038) | (0.031) | (0.031) | (0.089) | (0.077) | (0.076) |
| *N_ques_followed* | 0.0841*** | 0.0065 | −0.0036 | 0.1485*** | 0.0947** | 0.0655** | 0.0864*** |
| | (0.010) | (0.015) | (0.013) | (0.011) | (0.038) | (0.032) | (0.033) |
| *Constant* | −0.2673 | 1.3700 | 0.9644 | −0.0334 | 3.4807** | 3.6301*** | 2.9596** |
| | (0.302) | (0.849) | (0.648) | (0.296) | (1.584) | (1.365) | (1.347) |
| Observations | 30,272 | 8,260 | 8,260 | 30,272 | 9,550 | 9,550 | 9,550 |
| User fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted $R^2$ | 0.774 | 0.684 | 0.869 | 0.686 | 0.616 | 0.513 | 0.594 |
| Number of users | 1,892 | 1,128 | 1,128 | 1,892 | 1,356 | 1,356 | 1,356 |

*Note.* Robust standard errors clustered by user are in parentheses.
***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

*difficult to understand"*; $\alpha = 0.7$) and three items from Fang et al. (2012) to measure cognitive processing load (*"I expended a lot of mental energy reading the presented content"*, *"When reading the presented content, my information load was high"*, *"Reading the presented content requires a high level of information processing ability"*; $\alpha = 0.9$). We also self-created the following items: One item to measure mindset switching (*"The presentation order of the above content requires me to frequently switch my way of thinking while reading"*), two items to measure willingness to engage (*"As a user of this platform, I am willing to like/comment on the content I just read"*; $\alpha = 0.6$), and one item to measure willingness to contribute (*"As a user of this platform, I am willing to contribute my own content to this platform"*). All the measures were administrated in Chinese (see Online Appendix E for the Chinese versions of measurement items) and used a seven-point Likert scale (*1 = strongly disagree*, *7 = strongly agree*). Finally, subjects provided demographic information (e.g., gender, frequency of using Q&A sites).

Two participants did not pass simple attention checks and were excluded from analyses. In the attention checks, we asked whether there was video content (the true answer is no) and whether there was content about nonfungible token (the true answer is yes). Note that the attention-checking questions were administrated after subjects answered items related to their information processing experiences in order to prevent answering these questions from affecting reporting of their true experiences. Subjects need to answer all questions correctly to pass the attention check. In the end, we have 29 (28) valid responses for the Separated (Mixed) condition.

Table 12 presents the results. In Panel (a), we confirm the success of randomization: the subjects in the two conditions did not differ significantly on gender, or frequency of using online Q&A sites. In Panel (b), we find that subjects in the Mixed condition, relative to those in the Separated condition, reported less fluency in information processing experiences and higher cognitive load while processing the presented information, but these two effects are only close to statistically significant ($p = 0.12$ and $p = 0.13$, respectively). However, subjects reported significantly higher frequency of switching mindsets while reading the content ($p < 0.01$). Moreover, based on subjects' reported willingness to engage and contribute, mixing *answers* and *ideas* also significantly negatively affects users' engagement incentives ($p < 0.10$) and contribution incentives ($p < 0.10$).

Overall, the results from the laboratory experiment corroborate our findings based on the empirical data—mixed presentation of expert knowledge content and social posts could decrease users' willingness to engage (and also incentives to contribute). Moreover, it also offers some support to the theorization that mindset switching is the underlying mechanism since mixed presentation also increases subjects' reported frequency of mindset switching at the same time.

### 6.2. Content Contribution

In Section 5.3, we found that integrating social posts (*ideas*) with expert knowledge content (*answers*) decreased the contribution of both types of content, though more so for *ideas* than for *answers*. Now, we test two pieces of our theorization regarding the negative effects of integration on contribution: 1) integration caused a reduction in social recognition benefits (as users engaged less with contributed content), and 2) integration exacerbated the *multiple audience problem* (such that users worried that posting *ideas* would dilute their professional image). Considering the difficulty of simulating a context wherein users receive varied social recognition benefits and have varied image concerns in a laboratory experiment, we next examine these two mechanisms by conducting additional econometric tests based on the field data.

First, regarding social recognition benefits, we find that a user's content contribution in the current period is positively predicted by volumes of engagements (e.g., comments, voteups, thanks, etc.) they received on content in the previous period. The results are consistent with our argument about social recognition as an incentive to contribute. Considering this is an already-established effect in the extant literature (e.g., von Krogh et al. 2012, Chen et al. 2018, Kuang et al. 2019)

**Table 12.** Comparisons of Means Between Conditions in the Experiment

| Variables | Separated N | Separated Mean [95% CI] | Mixed N | Mixed Mean [95% CI] | $t$-test of difference (Separated - Mixed) |
|---|---|---|---|---|---|
| (a) Demographics: | | | | | |
| Female | 29 | 0.83 [0.68, 0.97] | 28 | 0.75 [0.58, 0.92] | $t$ = 0.71 ($p$ = 0.48) |
| Frequency of using Q&A site | 29 | 5.45 [4.91, 5.98] | 28 | 5.25 [4.71, 5.79] | $t$ = 0.53 ($p$ = 0.60) |
| (b) Dependent measures: | | | | | |
| Processing fluency | 29 | 3.97 [3.42, 4.51] | 28 | 3.43 [2.98, 3.88] | $t$ = 1.56 ($p$ = 0.12) |
| Cognitive processing load | 29 | 4.99 [4.46, 5.52] | 28 | 5.49 [5.09, 5.89] | $t$ = −1.53 ($p$ = 0.13) |
| Mindset switching | 29 | 4.52 [3.85, 5.19] | 28 | 6.07 [5.71, 6.44] | $t$ = −4.13 ($p$ = 0.00) |
| Willingness to engage | 29 | 4.64 [4.20, 5.08] | 28 | 4.02 [3.51, 4.53] | $t$ = 1.89 ($p$ = 0.06) |
| Willingness to contribute | 29 | 4.72 [4.13, 5.32] | 28 | 4.00 [3.41, 4.59] | $t$ = 1.78 ($p$ = 0.08) |

and also due to limited space in the main text, we report these results in Tables F1–F5, Online Appendix F.

Second, to evaluate the multiple audience problem, we parsed the text of the *ideas* in our sample to classify them into two categories: closely related to *answers* (such that the *idea* text contains a link to an *answer*) and unrelated to *answers*. We argue that image-dilution concerns should be less salient for *ideas* that are closely related to *answers* since the content of closely related *ideas* is fairly more consistent with a professional expert image. Following this logic, we expect that integration led to a smaller decrease in the contribution of *ideas* that are closely related (versus unrelated) to *answers*.

We estimate Equations (1) and (2) on the separate subsamples of *ideas* that are closely related to *answers* (Table 13) and *ideas* that are unrelated to *answers* (Table 14). In Table 13, the estimation results in Columns (1) and (2) indicate that integration led to an average decrease of 5% in the contribution of *ideas* that are closely related to *answers*. The estimated effect is smaller (3%) when the maximum polynomial order is 2 and is insignificant when the maximum polynomial order is 3. In Table 14, the estimation results in Columns (1) and (2) indicate that integration led to an average decrease of 15% in the contribution of *ideas* that are unrelated to *answers*—three times the size of the effect on *ideas* that are closely related to *answers*. The estimated effects remain significant but smaller when the maximum polynomial order is 2 (8% decrease) and 3 (12% decrease). The results support our theorization based on image dilution concerns.

In Online Appendix F, we use an alternative approach to assess the similarity between *ideas* and *answers*. Specifically, we use a neural-network-based text analysis method called Doc2Vec to compute the semantic similarity between *ideas* and *answers*. Doc2Vec generates fixed-length feature representation from variable-length pieces of text, such as sentences, paragraphs, and documents, with good performance (Le and Mikolov 2014, Kim et al. 2019). This ability fits our data well because the lengths of *ideas* and *answers* vary considerably. Specifically, for each *idea* posted by user *i*, we calculate its similarity with each *answer* written by user *i*, and we keep the maximum similarity score as the score of this *idea*. We split the sample of *ideas* at the 75th percentile of the similarity scores to create two subsamples of *ideas*—

**Table 13.** Effects of Integration on Contribution of *Ideas* That Are Closely Related to *Answers*

| Variables | Polynomial Order 0 (1) | Polynomial Order 1 (2) | Polynomial Order 2 (3) | Polynomial Order 3 (4) |
|---|---|---|---|---|
| *Integration* | −0.0497*** | −0.0498*** | −0.0357** | −0.0148 |
| | (0.012) | (0.012) | (0.017) | (0.022) |
| *Duration* | | −0.0004 | 0.0004 | 0.0078 |
| | | (0.001) | (0.003) | (0.006) |
| *Duration*$^2$ | | | −0.0002 | −0.0021* |
| | | | (0.000) | (0.001) |
| *Duration*$^3$ | | | | 0.0001 |
| | | | | (0.000) |
| *Integration* × *Duration*$^2$ | | | 0.0003 | 0.0021 |
| | | | (0.000) | (0.002) |
| *Integration* × *Duration*$^3$ | | | | −0.0001 |
| | | | | (0.000) |
| *Age* | 0.2127* | 0.2091* | −0.0509 | −0.6711 |
| | (0.114) | (0.114) | (0.277) | (0.561) |
| *Age*$^2$ | −0.0008 | −0.0004 | −0.0003 | −0.0001 |
| | (0.020) | (0.020) | (0.020) | (0.020) |
| *N_ans_liked* | 0.0710*** | 0.0710*** | 0.0707*** | 0.0706*** |
| | (0.007) | (0.007) | (0.007) | (0.007) |
| *N_ques_asked* | 0.0117 | 0.0115 | 0.0115 | 0.0115 |
| | (0.023) | (0.023) | (0.023) | (0.023) |
| *N_ques_followed* | 0.0401*** | 0.0401*** | 0.0400*** | 0.0399*** |
| | (0.009) | (0.009) | (0.009) | (0.009) |
| *Constant* | −0.3897** | −0.3809** | 0.2121 | 1.6278 |
| | (0.171) | (0.170) | (0.610) | (1.261) |
| Observations | 15,368 | 15,368 | 15,368 | 15,368 |
| User fixed effects | Yes | Yes | Yes | Yes |
| Adjusted $R^2$ | 0.523 | 0.523 | 0.523 | 0.523 |
| Number of users | 1,257 | 1,257 | 1,257 | 1,257 |

*Notes.* The interaction term *Integration* × *Duration* is omitted because it is collinear with *Age*. Robust standard errors clustered by user are in parentheses.
$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

**Table 14.** Effects of Integration on Contribution of *Ideas* That Are Unrelated to *Answers*

| Variables | Polynomial Order 0 (1) | Polynomial Order 1 (2) | Polynomial Order 2 (3) | Polynomial Order 3 (4) |
|---|---|---|---|---|
| *Integration* | −0.1581*** | −0.1578*** | −0.0796*** | −0.1266*** |
| | (0.019) | (0.019) | (0.022) | (0.029) |
| *Duration* | | 0.0011 | 0.0048 | −0.0226** |
| | | (0.001) | (0.004) | (0.009) |
| *Duration*$^2$ | | | −0.0011*** | 0.0045** |
| | | | (0.000) | (0.002) |
| *Duration*$^3$ | | | | −0.0002*** |
| | | | | (0.000) |
| *Integration* × *Duration*$^2$ | | | 0.0018*** | −0.0009 |
| | | | (0.000) | (0.003) |
| *Integration* × *Duration*$^3$ | | | | 0.0001 |
| | | | | (0.000) |
| *Age* | 0.5333** | 0.5432** | −0.9092** | 0.2238 |
| | (0.251) | (0.251) | (0.437) | (0.890) |
| *Age*$^2$ | −0.0779* | −0.0791* | −0.0786* | −0.0792* |
| | (0.046) | (0.046) | (0.046) | (0.046) |
| *N_ans_liked* | 0.1039*** | 0.1038*** | 0.1022*** | 0.1024*** |
| | (0.010) | (0.010) | (0.010) | (0.010) |
| *N_ques_asked* | 0.0998*** | 0.1003*** | 0.1001*** | 0.1002*** |
| | (0.030) | (0.030) | (0.030) | (0.030) |
| *N_ques_followed* | 0.0785*** | 0.0786*** | 0.0783*** | 0.0785*** |
| | (0.012) | (0.012) | (0.012) | (0.012) |
| *Constant* | 0.4132 | 0.3887 | 3.7037*** | 1.1416 |
| | (0.341) | (0.342) | (0.868) | (1.998) |
| Observations | 15,368 | 15,368 | 15,368 | 15,368 |
| User fixed effects | Yes | Yes | Yes | Yes |
| Adjusted $R^2$ | 0.657 | 0.657 | 0.658 | 0.658 |
| Number of users | 1,257 | 1,257 | 1,257 | 1,257 |

*Notes.* The interaction term *Integration* × *Duration* is omitted because it is collinear with *Age*. Robust standard errors clustered by user are in parentheses.

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

closely related versus unrelated to *answers*. (We choose a relatively large splitting cutoff to ensure we can isolate *ideas* that are indeed closely related to *answers* from other *ideas*.) We repeat the analyses reported in Tables 13 and 14. The results, reported in Tables F6 and F7, Online Appendix F, are qualitatively consistent.

Moreover, we follow Xu et al. (2020) by adopting a modified version of the difference-in-differences approach by comparing users' contributions of *ideas* related to versus unrelated to *answers* before and after the information feed integration. The results are highly consistent with the results reported in Tables 13 and 14 (see Table F8, Online Appendix F)

We also investigated the role of image dilution concerns by testing whether the user's investment in their professional image moderates the effect of integration on the contribution of *ideas*. If information feed integration heightens image dilution concerns, as we have theorized, then the decrease in the contribution of *ideas* should be larger for users who are more invested in their professional image. Following this reasoning, we classified users into two groups based on the number of *answers* certified by the platform. (The platform certifies and awards badges to popular, high-quality *answers*.) We reason that users who obtained more badges were more invested in their professional image. We created a dummy, *High_ans_certified*, that equals 1 (0) for users with an above-median (below-median) number of certified *answers*. We added the interaction term, *High_ans_certified* × *Integration*, to Equations (1) and (2).

In Table 15, the estimated coefficient of the interaction term, *High_ans_certified* × *Integration*, is significantly negative in all four specifications, consistent with our theorization based on concerns about diluting one's professional image.

We repeat the analysis with a different approach to classifying users as more or less invested in their professional image. Zhihu hosts paid knowledge-sharing activities (called *Live Talks*) as well as free Q&A posts, and monetary rewards are known to be associated with the professional expert image (Wang et al. 2022). We add the interaction term *LiveSpeaker* × *Integration* to Equations (1) and (2), where *LiveSpeaker* is a dummy that equals 1 for users who gave one or more live talks, and 0 otherwise. The results are reported in Table F9, Online Appendix F. As expected, we find that integration led to a significantly larger decrease in the

**Table 15.** Effects of Integration on Contributions of *Ideas*, Contingent on Contributing Many Certified *Answers*

| Variables | Polynomial Order 0 (1) | Polynomial Order 1 (2) | Polynomial Order 2 (3) | Polynomial Order 3 (4) |
|---|---|---|---|---|
| *Integration* | −0.0603*** | −0.0615*** | −0.0096 | −0.0190 |
| | (0.011) | (0.011) | (0.011) | (0.013) |
| *High_ans_certified* × *Integration* | −0.0498*** | −0.0501*** | −0.0501*** | −0.0501*** |
| | (0.015) | (0.015) | (0.015) | (0.015) |
| *Duration* | | −0.0014*** | 0.0002 | −0.0115*** |
| | | (0.001) | (0.002) | (0.003) |
| $Duration^2$ | | | −0.0007*** | 0.0012 |
| | | | (0.000) | (0.001) |
| $Duration^3$ | | | | −0.0001** |
| | | | | (0.000) |
| *Integration* × $Duration^2$ | | | 0.0012*** | 0.0012 |
| | | | (0.000) | (0.001) |
| *Integration* × $Duration^3$ | | | | −0.0000 |
| | | | | (0.000) |
| *Age* | 0.1505 | 0.1433 | −0.8098*** | −0.7136** |
| | (0.115) | (0.114) | (0.195) | (0.353) |
| $Age^2$ | −0.0019 | −0.0004 | −0.0003 | −0.0003 |
| | (0.020) | (0.020) | (0.020) | (0.020) |
| *N_ans_liked* | 0.0910*** | 0.0910*** | 0.0906*** | 0.0907*** |
| | (0.006) | (0.006) | (0.006) | (0.006) |
| *N_ques_asked* | 0.1050*** | 0.1046*** | 0.1041*** | 0.1040*** |
| | (0.027) | (0.027) | (0.027) | (0.027) |
| *N_ques_followed* | 0.0864*** | 0.0864*** | 0.0864*** | 0.0864*** |
| | (0.008) | (0.008) | (0.008) | (0.008) |
| *Constant* | −0.0734 | −0.0549 | 2.1195*** | 1.9170** |
| | (0.157) | (0.157) | (0.384) | (0.783) |
| Observations | 58,240 | 58,240 | 58,240 | 58,240 |
| User fixed effects | Yes | Yes | Yes | Yes |
| Adjusted $R^2$ | 0.730 | 0.730 | 0.730 | 0.730 |
| Number of users | 1,820 | 1,820 | 1,820 | 1,820 |

*Notes.* The interaction term *Integration* × *Duration* is omitted because it is collinear with *Age*. *High_ans_certified* equals 1 (0) for users with an above-median (below-median) number of certified *answers*. The singular term of *High_ans_certified* is omitted because it collinear with the user fixed-effects. Robust standard errors clustered by user are in parentheses.

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

contribution of *ideas* among users who conducted at least one live talk on Zhihu.

## 7. Discussion and Conclusion

### 7.1. Implications

Online UGC is increasingly diverse due to developments in Internet technologies, and it also is increasingly cheap to host diverse UGC—so it is tempting for platform owners to curate diverse UGC on the same platform for the purpose of increasing engagement and content contributions. Indeed, users often wish to present multiple selves and contribute diverse content on the same platform (Schau and Gilly 2003, Belk, 2013).

In this research, however, we show that the effectiveness of a diversity-oriented strategy might be contingent on a basic but critical factor: the presentation format of heterogeneous content types. Our study focuses on a simple but common design choice: whether to present diverse UGC in one information feed or in separate feeds. In a natural experiment involving integration of social content into the main feed of expert knowledge content, we find that integration significantly decreased user engagement with both content types and also decreased the volumes of contributions. It seems that the benefits of implementing an informal social space on knowledge-sharing platforms are contingent on keeping the social content separate from the main knowledge content.

Our findings are consistent with established theories of information processing, particularly mindset theory (French II 2016) and schema theory (Bartlett 1932, Axelrod 1973). The theories posit that people process different types of information and knowledge with different mindsets or schemas, so the integration of content that fits/requires different mindsets/schemas can increase cognitive load and hurt the information processing experience (Meyers-Levy and Tybout 1989, Aggarwal and McGill 2007, Hamilton et al. 2011, Yan et al. 2018). To the best of our knowledge, we are the first to apply these information processing theories to understand how information presentation formats affect online users' content engagement and contribution behaviors.

Our results also show that changes in users' contribution behaviors seem to involve concerns about diluting their professional image. We find that users who are more invested in their professional image on the platform are more likely to refrain from contributing social content after its integration with expert knowledge content. Their reluctance likely involves the *multiple audience problem* (Fleming et al. 1990, Schlosser 2005, Lee et al. 2015, Gil-Lopez et al. 2018); integration heightened users' concerns that posting social content would dilute their professional image, so the integration of disparate audiences led to the "crowding out" of some contribution motives due to the *lowest common denominator effect* (Hogan 2010, Marder et al. 2016). A crowding-out effect suggests that the value of an informal, virtual "third place" (Chen et al. 2021) is inhibited when the "third place" intrudes into and conflicts with a professional sphere.

The results have practical implications for any platform that hosts heterogeneous types of content. In general, platform owners should consider the extent to which different types of content are (in)congruous in terms of information processing (from the reader's perspective) as well as virtual self-presentation (from the contributor's perspective). With this awareness, platform owners can make strategic decisions about displaying heterogeneous content. If the types of content are fairly incongruous, then we would recommend separating the content into different information feeds.

Then, if a platform must unify information feeds, platform operators may be able to mitigate the negative consequences of integration by aiding users to acquire the type(s) of information they seek and by reducing contributors' concerns about the multiple audience problem. For users, a "filtering" function might help them focus on one type of content at a time. Closely related content could be clustered within the feed (instead of a strictly chronological display) to reduce the frequency with which users must switch mindsets as they scroll down. In fact, our online experiment (Section 6.1) confirmed that users who browsed heterogeneous content in separate clusters had better information processing experiences than users for whom heterogeneous content was mixed. For contributors, platform operators may be able to alleviate a crowding-out effect by offering customized settings with which contributors can choose whether their social content should appear in the integrated information feed. For example, many popular social media platforms, including Facebook, Sina Weibo, and WeChat Moments, enable users to choose which audiences (e.g., friends, colleges, relatives, and fans) have access to the posted content.

Interestingly, the relatively new Instagram feature, *Stories*, allows users to create collections of pictures, either in photo albums or videos, to tell "stories" about themselves. *Stories* appear at the top of the main feed, so a user can browse pictures in the main feed, tap on a story to open it in a separate space, and then return to the main feed. Instagram's approach represents a compromise between fully separating and fully mixing different content types. Based on our results, we would speculate that the diversity of content is engaging and satisfying, and the (partial) separation of the content types reduces users' cognitive load and improves the information processing experience.

### 7.2. Conclusion and Future Research

While many online communities host multiple types of content to satisfy users' heterogeneous interests, little is known about how the presentation format of different types of content affects user engagement with and contribution of content. To the best of our knowledge, we are the first empirical study to fill this gap. We exploit a natural experiment on Zhihu, the leading Chinese online Q&A platform, in which social content was moved from a separate feed into the main feed, where it was integrated with expert knowledge content.

We find that information feed integration significantly decreased user engagement with and contribution of both social and expert content. The results are consistent with the hypothesis that merging heterogeneous content leads users to engage less because the juxtaposition of content with incongruent mindsets reduces information processing fluency. Then, content generation decreases, both because of the decrease in engagement (weaker social recognition incentives) and because integration heightened concerns that posting social content would dilute the contributor's professional image. Our findings have important theoretical and practical implications for any platform that hosts heterogeneous content.

We wish to highlight several limitations of our research that warrant further investigation. First, the natural experiment we exploit involved a relatively simple change in the presentation format: from entirely separate information feeds to a fully integrated one in which all content is presented chronologically. In practice, platforms can make more nuanced design choices, for example, by including a filter function or clustering information by type within the same feed. It would advance our theoretical understanding of the mechanisms, and would provide better guidance for platform managers, to conduct similar analyses in settings with nuanced integration features.

Second, the data we collected only includes information about users' engaging and contributing behaviors; we did not have information about users' other activities such as log-ins and content reading. Further research could investigate whether and how these outcomes are affected by information feed integration, which could provide additional insights for the platform managers. Moreover, if detailed information about when the engagements occurred on the content is available, it is worthwhile to explore the impacts of integration on the evolution of user engagements. This could offer a more nuanced perspective on how integration affects the dynamics of user engagement over time.

Third, we did not include much textual analysis, but sophisticated text analysis techniques could be used to investigate how integration affects users' linguistic choices (e.g., sentiment, neutrality) in different content types, which can help reveal potentially more nuanced changes in users' contribution patterns.

Fourth, our distinction between social content and expert knowledge content is most relevant to knowledge-sharing platforms. We hope future research will investigate the effects of presentation format choices involving content that is heterogeneous in other ways (e.g., text versus image versus video).

Finally, it would be useful to understand whether and how, if yes, information feed integration could trigger other social mechanisms such as altruism and reciprocity that are common in online social communities, provided that such granular data are available.

## Acknowledgments

The authors thank the senior editor, associate editor, and four anonymous reviewers for their helpful comments. The authors also thank participants at the 2021 Workshop on Information Systems and Economics, 2021 INFORMS Annual Meeting, and Jing Wang, Tat Koon Koh, and Jinyang Zheng for their helpful suggestions.

## Endnotes

[1] See https://www.quora.com/Whats-a-Quora-post (accessed March 31, 2023).

[2] This new function was first implemented in Zhihu's mobile application and then rolled out to its website. See https://tech.huanqiu.com/article/9CaKrnK4TD5 (in Chinese, accessed March 31, 2023).

[3] The different natures of the two types of content also lead to differences on specific functions of the two content types. Prominently, *ideas* content supports the common hashtag functionality but *answers* content does not.

[4] On Zhihu, *articles* are another form of expert knowledge content users contribute on certain topics. Different from *answers*, *articles* need no preexisting questions.

[5] See https://baijiahao.baidu.com/s?id=1686394915709760705 (accessed March 31, 2023).

[6] The percentage change is calculated as $(e^{-0.2693} - 1) \times 100\% = -24\%$; the negative sign means a decrease.

## References

Aggarwal P, McGill AL (2007) Is that car smiling at me? Schema congruity as a basis for evaluating anthropomorphized products. *J. Consumer Res.* 34(4):468–479.

Agnew JR, Szykman LR (2005) Asset allocation and information overload: The influence of information display, asset choice, and investor experience. *J. Behav. Finance* 6(2):57–70.

Alba JW, Hasher L (1983) Is memory schematic? *Psych. Bull.* 93(2): 203–231.

Anderson ML (2014) Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion. *Amer. Econom. Rev.* 104(9): 2763–2796.

Ask K, Granhag PA (2005) Motivational sources of confirmation bias in criminal investigations: The need for cognitive closure. *J. Investigative Psych. Offender Profiling* 2(1):43–63.

Auffhammer M, Kellogg R (2011) Clearing the air? The effects of gasoline content regulation on air quality. *Amer. Econom. Rev.* 101(6):2687–2722.

Axelrod R (1973) Schema theory: An information processing model of perception and cognition. *Amer. Political Sci. Rev.* 67(4):1248–1266.

Baek J, Shore J (2020) Forum size and content contribution per person: A field experiment. *Management Sci.* 66(12):5485–6064.

Bapna R, Ramaprasad J, Umyarov A (2018) Monetizing freemium communities: Does paying for premium increase social engagement? *MIS Quart.* 42(3):719–735.

Bartlett F (1932) *Remembering: A Study in Experimental and Social Psychology* (Cambridge University Press, London).

Belk RW (2013) Extended self in a digital world. *J. Consumer Res.* 40(3):477–500.

Bettman JR, Kakkar P (1977) Effects of information presentation format on consumer information acquisition strategies. *J. Consumer Res.* 3(4):233–240.

Bettman JR, Johnson EJ, Payne JW (1991) Consumer decision making. Robertson TS, Kassarjian HH, eds. *Handbook of Consumer Behaviour* (Prentice-Hall, Englewood Cliffs, NJ) 50–84.

Bonaccorsi A, Giannangeli S, Rossi C (2006) Entry strategies under competing standards: Hybrid business models in the open source software industry. *Management Sci.* 52(7):1085–1098.

Bullingham L, Vasconcelos AC (2013) 'The presentation of self in the online world': Goffman and the study of online identities. *J. Inform. Sci.* 39(1):101–112.

Burtch G, Hong Y, Bapna R, Griskevicius V (2017) Stimulating online reviews by combining financial incentives and social norms. *Management Sci.* 64(5):2065–2082.

Chen W, Wei X, Zhu K (2018) Engaging voluntary contributions in online communities: A hidden Markov model. *MIS Quart.* 42(1):83–100.

Chen X, Forman C, Kummer M (2021) Chat more and contribute better: An empirical study of a knowledge-sharing community. Preprint, https://ssrn.com/abstract=3906520.

Chen Y, Cheng HK, Liu Y, Pu J, Qiu L, Wang N (2022) Knowledge-sharing ties and equivalence in corporate online communities: A novel source to understand voluntary turnover. *Production Oper. Management* 31(10):3896–3913.

Davis LW (2008) The effect of driving restrictions on air quality in Mexico City. *J. Political Econom.* 116(1):38–81.

Engel JF, Kollat DT, Blackwell RD (1968) *Consumer Behavior* (Holt, Rinehart, and Winston Place, New York).

Esfandiari M, Borromeo RM, Nikookar S, Sakharkar P, Amer-Yahia S, Roy SB (2021) Multi-session diversity to improve user satisfaction in web applications. *Proc. Web Conf. 2021* (Association for Computing Machinery, New York), 1928–1936.

Fang X, Hu PJ-H, Chau M, Hu H-F, Yang Z, Liu Sheng OR (2012) A data-driven approach to measure website navigability. *J. Management Inform. Systems* 29(2):173–212.

Festinger L (1957) *A Theory of Cognitive Dissonance*, vol. 2 (Stanford University Press, Redwood City, CA).

Fleming JH, Darley JM, Hilton JL, Kojetin BA (1990) Multiple audience problem: A strategic communication perspective on social perception. *J. Personality Soc. Psych.* 58(4):593–609.

French II RP (2016) The fuzziness of mindsets: Divergent conceptualizations and characterizations of mindset theory and praxis. *Internat. J. Organ. Anal.* 24(4):673–691.

Gil-Lopez T, Shen C, Benefield GA, Palomares NA, Kosinski M, Stillwell D (2018) One size fits all: Context collapse, self-presentation strategies and language styles on Facebook. *J. Comput. Mediat. Comm.* 23(3):127–145.

Goes PB, Guo C, Lin M (2016) Do incentive hierarchies induce user effort? Evidence from an online knowledge exchange. *Inform. Systems Res.* 27(3):497–516.

Goes PB, Lin M, Au Yeung CM (2014) "Popularity effect" in user-generated content: Evidence from online product reviews. *Inform. Systems Res.* 25(2):222–238.

Goffman E (1978) *The Presentation of Self in Everyday Life*, vol. 21 (Harmondsworth, London).

Goodman LA (1961) Snowball sampling. *Ann. Math. Statist.* 32(1): 148–170.

Hamilton R, Vohs KD, Sellier A-L, Meyvis T (2011) Being of two minds: Switching mindsets exhausts self-regulatory resources. *Organ. Behav. Human Decision Processes* 115(1):13–24.

Hausman C, Rapson DS (2018) Regression discontinuity in time: Considerations for empirical applications. *Annual Rev. Resour. Econom.* 10:533–552.

Hogan B (2010) The presentation of self in the age of social media: Distinguishing performances and exhibitions online. *Bull. Sci. Tech. Soc.* 30(6):377–386.

Huang N, Hong Y, Burtch G (2017) Social network integration and user content generation: Evidence from natural experiments. *MIS Quart.* 41(4):1035–1058.

Ives B, Hamilton S, Davis GB (1980) A framework for research in computer-based management information systems. *Management Sci.* 26(9):910–934.

Kahneman D (1973) *Attention and Effort* (Prentice-Hall, Englewood Cliffs, NJ).

Kahneman D (2011) *Thinking, Fast and Slow* (Macmillan, New York).

Kim A, Dennis AR (2019) Says who? The effects of presentation format and source rating on fake news in social media. *MIS Quart.* 43(3):1025–1039.

Kim D, Seo D, Cho S, Kang P (2019) Multi-co-training for document classification using various document representations: TF–IDF, LDA, and Doc2Vec. *Inform. Sci.* 477:15–29.

Kruglanski AW, Jasko K, Milyavsky M, Chernikova M, Webber D, Pierro A, Di Santo D (2018) Cognitive consistency theory in social psychology: A paradigm reconsidered. *Psych. Inquiry* 29(2): 45–59.

Kuang L, Huang N, Hong Y, Yan Z (2019) Spillover effects of financial incentives on non-incentivized user engagement: Evidence from an online knowledge exchange platform. *J. Management Inform. Systems* 36(1):289–320.

Le Q, Mikolov T (2014) Distributed representations of sentences and documents. *Internat. Conf. Machine Learn.* (PMLR, New York), 1188–1196.

Lee AY, Keller PA, Sternthal B (2010) Value from regulatory construal fit: The persuasive impact of fit between consumer goals and message concreteness. *J. Consumer Res.* 36(5):735–747.

Lee DS, Lemieux T (2010) Regression discontinuity designs in economics. *J. Econom. Lit.* 48(2):281–355.

Lee S-Y, Qiu L, Whinston A (2018) Sentiment manipulation in online platforms: An analysis of movie tweets. *Production Oper. Management* 27(3):393–416.

Lee Y-J, Hosanagar K, Tan Y (2015) Do I follow my friends or the crowd? Information cascades in online movie ratings. *Management Sci.* 61(9):2241–2258.

Lou J, Fang Y, Lim KH, Peng JZ (2013) Contributing high quantity and quality knowledge to online Q&A communities. *J. Amer. Soc. Inform. Sci. Tech.* 64(2):356–371.

Ma M, Agarwal R (2007) Through a glass darkly: Information technology design, identity verification, and knowledge contribution in online communities. *Inform. Systems Res.* 18(1):42–67.

Maines LA, McDaniel LS (2000) Effects of comprehensive-income characteristics on nonprofessional investors' judgments: The role of financial-statement presentation format. *Accounting Rev.* 75(2):179–207.

Marder B, Joinson A, Shankar A, Thirlaway K (2016) Strength matters: Self-presentation to the strongest audience rather than lowest common denominator when faced with multiple audiences in social network sites. *Comput. Human Behav.* 61:56–62.

Marwick AE, Boyd D (2011) I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media Soc.* 13(1):114–133.

Meyers-Levy J, Tybout AM (1989) Schema congruity as a basis for product evaluation. *J. Consumer Res.* 16(1):39–54.

Moe WW, Trusov M (2011) The value of social dynamics in online product ratings forums. *J. Marketing Res.* 48(3):444–456.

Nisbett RE, Zukier H, Lemley RE (1981) The dilution effect: Non-diagnostic information weakens the implications of diagnostic information. *Cognit. Psych.* 13(2):248–277.

Palmer JW (2002) Website usability, design, and performance metrics. *Inform. Systems Res.* 13(2):151–167.

Pu J, Chen Y, Qiu L, Cheng HK (2020) Does identity disclosure help or hurt user content generation? Social presence, inhibition, and displacement effects. *Inform. Systems Res.* 31(2):297–322.

Pullig C, Simmons CJ, Netemeyer RG (2015) Brand dilution: When do new brands hurt existing brands? *J. Marketing* 70(2): 52–66.

Qiu L, Kumar S (2017) Understanding voluntary knowledge provision and content contribution through a social-media-based prediction market: A field experiment. *Inform. Systems Res.* 28(3): 529–546.

Schau HJ, Gilly MC (2003) We are what we post? Self-presentation in personal web space. *J. Consumer Res.* 30(3):385–404.

Schlosser AE (2005) Posting vs. lurking: Communicating in a multiple audience context. *J. Consumer Res.* 32(2):260–265.

Schulze C, Scholer L, Skiera B (2014) Not all fun and games: Viral marketing for utilitarian products. *J. Marketing* 78(1):1–19.

Shanteau J (1975) Averaging vs. multiplying combination rules of inference judgment. *Acta Psych. (Amsterdam)* 39:83–89.

Shen W, Hu YJ, Ulmer JR (2015) Competing for attention: An empirical study of online reviewers' strategic behavior. *MIS. Quart.* 39(3):683–696.

Shriver SK, Nair HS, Hofstetter R (2013) Social ties and user-generated content: Evidence from an online social network. *Management Sci.* 59(6):1425–1443.

Song T, Tang Q, Huang J (2019) Triadic closure, homophily, and reciprocation: An empirical investigation of social ties between content providers. *Inform. Systems Res.* 30(3):912–926.

Speier C, Valacich JS, Vessey I (1999) The influence of task interruption on individual decision making: An information overload perspective. *Decision Sci.* 30(2):337–360.

Speier C, Vessey I, Valacich JS (2003) The effects of interruptions, task complexity, and information presentation on computer-supported decision-making performance. *Decision Sci.* 34(4):771–797.

Toubia O, Stephen AT (2013) Intrinsic vs. image-related utility in social media: Why do people contribute content to Twitter? *Marketing Sci.* 32(3):365–531.

von Krogh G, Haefliger S, Spaeth S, Wallin MW (2012) Carrots and rainbows: Motivation and social practice in open source software development. *MIS Quart.* 36(2):649–676.

Wang J, Li G, Hui K-L (2022) Monetary incentives and knowledge spillover: Evidence from a natural experiment. *Management Sci.* 68(5):3549–3572.

Wasko MM, Faraj S (2005) Why should i share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quart.* 29(1):35–57.

Wheeler SC, Petty RE, Bizer GY (2005) Self-schema matching and attitude change: Situational and dispositional determinants of message elaboration. *J. Consumer Res.* 31(4):787–797.

Wooldridge JM (2010) *Econometric Analysis of Cross Section and Panel Data* (MIT Press, Cambridge, MA).

Wu L (2013) Social network effects on productivity and job security: Evidence from the adoption of a social networking tool. *Inform. Systems Res.* 24(1):30–51.

Wu Z, Zhou K, Liu Y, Zhang M, Ma S (2019) Does diversity affect user satisfaction in image search. *ACM Trans. Inform. Systems* 37(3):1–30.

Xu L, Nian T, Cabral L (2020) What makes geeks tick? A study of stack overflow careers. *Management Sci.* 66(2):587–604.

Yaari E, Baruchson-Arbib S, Bar-Ilan J (2011) Information quality assessment of community-generated content – a user study of Wikipedia. *J. Inform. Sci.* 37(5):487–498.

Yan J, Zhang N-N, Xu D-X (2018) Mindset switching increases the use of 'want-based' over 'should-based' behaviors. *PLoS One* 13(4):e0196269–e0196269.

Zhang XM, Zhu F (2011) Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *Amer. Econom. Rev.* 101(4):1601–1615.