



ISSN 1943-7544

Pacific Asia Journal of the Association for Information Systems

Research Paper

doi: 10.17705/1pais.13104

Volume 13, Issue 4 (2021)

Review Manipulation: Literature Review, and Future Research Agenda

Sana Ansari¹, Sumeet Gupta^{2,*}

¹Indian Institute of Management Raipur, India, sana.fpm2015@iimraipur.ac.in

²Indian Institute of Management Raipur, India, sumeetgupta@iimraipur.ac.in

Abstract

Background: The phenomenon of review manipulation and fake reviews has gained Information Systems (IS) scholars' attention during recent years. Scholarly research in this domain has delved into the causes and consequences of review manipulation. However, we find that the findings are diverse, and the studies do not portray a systematic approach. This study synthesizes the findings from a multidisciplinary perspective and presents an integrated framework to understand the mechanism of review manipulation.

Method: The study reviews 88 relevant articles on review manipulation spanning a decade and a half. We adopted an iterative coding approach to synthesizing the literature on concepts and categorized them independently into potential themes.

Results: We present an integrated framework that shows the linkages between the different themes, namely, the prevalence of manipulation, impact of manipulation, conditions and choice for manipulation decision, characteristics of fake reviews, models for detecting spam reviews, and strategies to deal with manipulation. We also present the characteristics of review manipulation and cover both operational and conceptual issues associated with the research on this topic.

Conclusions: Insights from the study will guide future research on review manipulation and fake reviews. The study presents a holistic view of the phenomenon of review manipulation. It informs various online platforms to address fake reviews towards building a healthy and sustainable environment.

Keywords: Fake Reviews, E-Commerce, Online Reviews, Integrated Framework.

Citation: Ansari, S., & Gupta, S. (2021). Review Manipulation: Literature Review, and Future Research Agenda. *Pacific Asia Journal of the Association for Information Systems*, 13(1), 97-121. <https://doi.org/10.17705/1pais.13104>
Copyright © Association for Information Systems.

Introduction

Studies on online reviews and word-of-mouth (e.g., Chevalier & Mayzlin, 2006; King et al., 2014; Moe & Schweidel, 2012) have found a positive relationship between average review score and product sales. Positive reviews improve product sales and business reputation, whereas negative reviews hurt sales and damage the business reputation. However, with online reviews now established as a powerful marketing tool, online sellers attempt to influence customers' purchase decisions by injecting fake reviews on the platform (John, 2019). Several reports have confirmed the ongoing practice of review manipulation – an act of injecting misleading reviews to influence customer decisions. According to an investigation by Daily Mail (Kelly, 2019), companies rely on an army of testers who purchase products on Amazon and give five-star feedback. Each tester receives a refund for the product purchased and a token fee of £13 for writing a favorable review. Another report highlights how fake five-star reviews flood Amazon's review system (Smithers, 2019). Recently, Amazon deleted around 20000 five-star reviews (mostly on Chinese products) given by seven of its top ten reviewers (Nikolic, 2020). These reviewers used to obtain free products in exchange for reviews. Similar reports from Yelp admitting a quarter of reviews on its site as fake and reviewers' confessions on Amazon (Chen, 2017) paint a grave picture of the online review systems' condition. Such evidence raise question about the integrity of reviews present online and the signal they send about the product quality. This study helps us understand the phenomena of review manipulation – how, when, and why a review fraud is likely to happen and the consequences of review manipulation, based on the studies conducted in this field.

The literature on review manipulation is diverse, and research questions vary substantially. Thus, deriving a meaningful conclusion is difficult. The topic initially received attention from computer science and later from other fields such as marketing, economics, and information systems. Computer science studies (e.g., Jindal & Liu, 2008; Jindal et al., 2010; Mukherjee et al., 2013a) focused on differentiating fake reviews from genuine ones and detecting fake reviews and reviewers. Studies in marketing (Mayzlin, 2006) and business management (Dellarocas, 2006) examined the conditions under which the online forums remain persuasive despite manipulation as well as the impact of manipulation on platforms'¹ informativeness. These studies (e.g., Hu et al., 2012; Mayzlin et al., 2014) primarily focused on economic aspects of review manipulation and contributed to the literature by examining the antecedents and motivations for online retailers to commit fraud. Research on review manipulation has also witnessed a fair amount of attention from researchers in information systems (IS) in recent years. However, the studies on manipulation in online reviews are fragmented and scattered in many journals across different fields. The extant scholarly reviews on review manipulation highlight single streams in the area of review manipulation. For example, Heydari et al. (2015) and Ngai et al. (2011) focus primarily on spam detection and its techniques. Similarly, Wu et al. (2020) focus on antecedents and consequences of fake reviews.

Therefore, in the present study, we examine the phenomenon of review manipulation holistically and from a multidisciplinary perspective. We aim to integrate the findings on this topic and understand where the phenomenon of review manipulation stands in the literature. Accordingly, we raise the following questions. (i) *What do we know about the field's primarily empirical evidence and challenges in this area of research?* And (ii) *What are the major research areas and future avenues in the field of review manipulation?* To answer these questions, we use a systematic approach for conducting the literature review as suggested by

¹ By platform, we mean websites. We will use platforms and websites interchangeably henceforth.

Schryen (2015). Based on his guidelines, we present a framework that helps classify and integrate empirical and conceptual work on review manipulation.

This study makes three interesting contributions to research. First, it enhances our understanding of the phenomenon of review manipulation. It elaborates on the characteristics of review manipulation and provides a comprehensive definition of the phenomenon. Second, it presents a synthetic view of the key research findings through an integrative framework. To do so, we develop a comprehensive classifying research framework that summarizes the relationships examined in the literature. Third, we propose an integrated framework that highlights the avenues for future research. Our findings suggest that research should adopt a multidisciplinary perspective for a holistic understanding of review manipulation mechanisms to foster research in information systems. The study further contributes to the practice by informing platforms, policymakers, and customers about recommendations in dealing with review manipulation.

We organize the rest of the study as follows. We present the literature review in the second section, where we define and discuss various aspects that makes review manipulation an exciting phenomenon to study. In the third section, we present the methodology adopted for conducting this literature review and a few descriptive findings. In the fourth section, we present our integrated framework and the themes ensuing from the framework. In the fifth section, we conclude this study by presenting research gaps and suggesting future research directions.

Literature Review

What is Review Manipulation?

Review manipulation is a global phenomenon (Hu, Bose et al., 2011; Hu et al., 2012; Lappas et al., 2016) and is present in some form on all platforms. Because of their global reach, their manipulation poses a challenge for market regulation (Luca & Zervas, 2016). Business managers engage in review manipulation because of the growing market pressure for improving ratings and reviews (Dellarocas, 2006; Gössling et al., 2019). Due to their being one-to-many communication (King et al., 2014; Litvin et al., 2008), they strongly influence customers' purchase decisions. Their manipulation combined with other Internet characteristics (anonymity, spatial separation, and low barriers to entry) is likely to leave consumers misguided and with reduced trust on the platform.

The literature on online reviews has explored their economic impact and examined the factors that make a review helpful. However, scholars have largely overlooked investigating review manipulation. Review manipulation emerged as a separate research field in the early 2000s (see Ferrara (2019) for the history of various kinds of digital spam). However, it gained prominence only during the last few years. Hu et al. (2012, p. 674) define review manipulation as "*vendors, publishers, writers, or any third-party consistently monitoring the online reviews and posting non-authentic online reviews on behalf of customers when needed, with the goal of boosting the sales of their products*" (p. 674). Lappas et al. (2016) added the dimensions of review valence and competition to the definition. According to them, manipulation involves injecting fake positive reviews about self or fake negative reviews about competitors by unscrupulous sellers. Kumar, Qiu, et al. (2018, p. 850) define review manipulation objective in terms of product popularity and describes the phenomenon as a "*firm's strategic use of fake accounts or the hiring of real individual accounts to post overly positive (biased) messages to boost product demand. The goal of manipulation is to purposefully influence perceptions of product popularity*" (p.850). Further, Gössling et al. (2018, p. 488) define manipulation as "*any attempt to deliberately control or influence online reputation, either with regard to one's own business or that of a competitor*" (p.488). Xu et al. (2020, p. 1) refer review manipulation as

“*generation of untruthful reviews of travel experiences*” (p.1). From various definitions, we can infer that review manipulation involves a deliberate and controlled injection of false information targeting several people at the same time.

To conceptually understand the review manipulation, we look at the characteristics of deception (Kumar & Shah, 2018; Xiao & Benbasat, 2011).

- Deception is an intentional or deliberate act and accomplished by manipulating information in some way (Xiao & Benbasat, 2011, p. 172).
- Deception has an instrumental end goal—that is, to create or maintain a belief in another that the communicator him/herself believes to be false (Xiao & Benbasat, 2011, p. 172).
- False information is classified based on its intent and knowledge. The creator of the false information creates information with or without an intent to mislead or deceive, or creates information knowingly or unknowingly to influence a reader’s opinion or decision (Kumar & Shah, 2018).

Considering these characteristics, we define review manipulation as “*a deliberate act of manipulation of online reviews by businesses (through addition, deletion, or exaggeration) to deceive customers by knowingly fostering incorrect information and inducing an action that the customer would unlikely take without the manipulation.*” Considering this definition, we do not consider those businesses or sellers that try to solicit honest reviews (positive or negative, irrespective of star ratings) by sending a gift note or free products to the customers under the purview of review manipulation. However, such activities for reviews might result in politeness or positive bias. But as the paper deals with review manipulation, this is out of the scope of our study.

What Makes an Online Review (or Message) Deceptive?

According to the Information manipulation theory (IMT) – which treats deception as a message level characteristic – messages function deceptively when they covertly violate the principles of conversational exchanges (McCornack, 1992, p. 1). These principles set expectations regarding the *quantity, quality, manner* of presentation, and *relevancy* of the information to the conversation. These four dimensions of manipulation, when varied, produce deceptive messages. Because the violations are not apparent to the readers (or listeners), they assume that the reviewer is adhering to conversation principles. Thus, the malpractice of review manipulation results from the manipulation of information by employing different strategies.

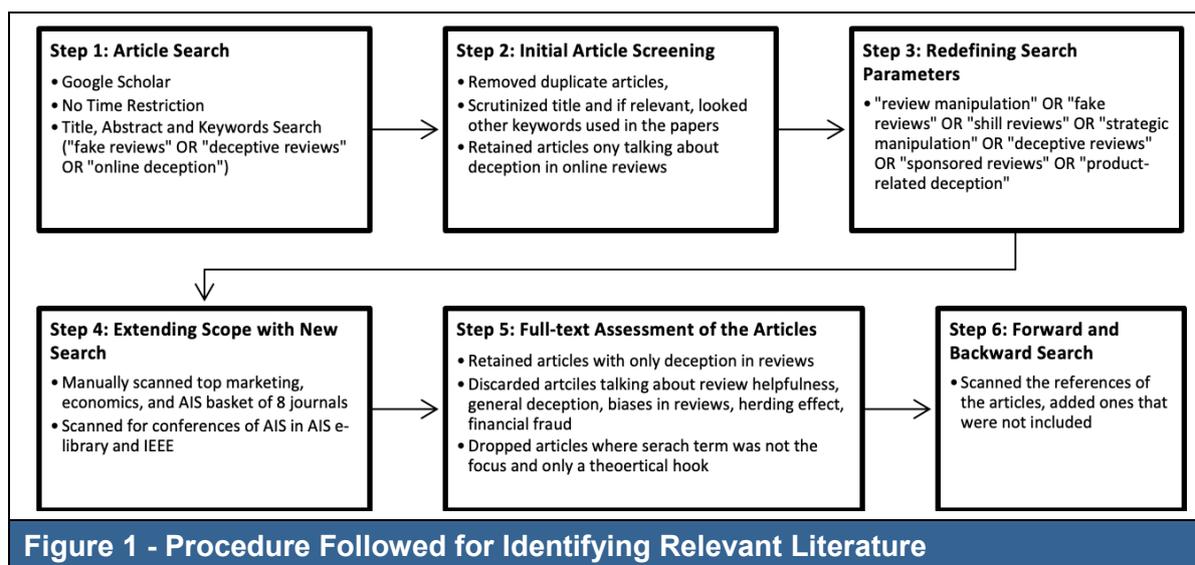
The primary reason a seller engages in review manipulation is to better rank on a platform (Lappas et al., 2016). Better ratings and ranking give a seller prominence and improve sales (Hu, Bose et al., 2011; Hu et al., 2012). The most common manipulation strategy is disguising. Online vendors, publishers, and authors impersonate as usual customers and engage in ballot stuffing using pseudo buyers (Hu, Bose et al., 2011; Munzel, 2016) who give high scores to themselves (Riazati et al., 2019). They also seek anonymous posting of biased reviews (Dellarocas, 2006; Mayzlin, 2006) from their friends and relatives (Hu, Bose et al., 2011; Hu et al., 2012), staff members on platforms and social networking sites (Gössling et al., 2018), through mechanical Turks (Luca & Zervas, 2016), or by partnering with the platforms (Gössling et al., 2018). Some sellers also appoint reputation management companies to monitor and manage the response to reviews on the platform (Munzel, 2016; Xu et al., 2020). Similar practices are also prevalent in the hospitality industry. These include negotiation and offering compensation to customers for removing negative reviews. Sellers also lure customers with gifts (in-cash or in-kind) in exchange for better reviews. Sellers also place negative reviews for competitors on other competing websites, which do not moderate the reviews on their platform (Gössling et al., 2018; Gössling et al., 2019). One of the common mechanisms through which sellers infiltrate into the review systems is WhatsApp, Facebook, and Telegram

groups. Sellers in these groups post images of the product for which they want a review. Interested group participants can directly message a seller and receive directions for posting the review. They then refund the participants with the amount of product and sometimes also provide a commission after the review posted goes live.

Platforms² also overtly favor review manipulation by providing incentives to customers for submitting a favorable review. However, they apparently try to address this growing concern about fake reviews by implementing various defense mechanisms, such as the verified buyer badge and the helpful vote. However, sellers found a way around these mechanisms also. For example, the platform adds a verified buyer badge to the reviews to signal that a genuine buyer has reviewed the product (or service). Online sellers purchase a competitor's product (or service) to earn the verified badge and then inject fake negative reviews (Lappas et al., 2016). The return policy on platforms makes it easier to inject self-praising reviews and negative reviews for competitors bearing almost no cost to the spam reviewers. Helpfulness count on platforms is also subject to manipulation. This mechanism allows other customers to rate a review as helpful or otherwise. Prior literature (Chua & Banerjee, 2015; Connors et al., 2011; Mudambi & Schuff, 2010) finds the positive significance of review helpfulness in customer's purchase decisions. However, sellers try to inflate the review helpfulness vote (Wan & Nakayama, 2014). That is why we find several products having many ratings and helpfulness votes compared to textual reviews.

Method and Descriptive Findings

To identify the relevant articles, we follow Webster and Watson's (Webster & Watson, 2002) suggestions and Schryen's (2015) guidelines on the literature search process. We focused on those articles that discussed manipulation or deception in online reviews. These articles spanned fifteen years (2006-2020). The search was performed in the year 2017 using the reference management tool (Mendeley) through its personalized recommendation feature, which kept us updated about any new articles published in the field. We conducted our search in six steps (Schryen, 2015, p. 12). Figure 1 describes the six-step procedure followed for the identification and assessment of papers in detail.



² Platforms running on marketplace model list several sellers which sell their products on the platform

First, we performed the topic search (including title, abstract, and keywords) using the terms “fake reviews” OR “online deception” OR “deceptive reviews” on Google Scholar. We did not restrict our search to any specific period. We selected relevant articles based on the title search. On examination, we found new terms used in the literature for deception in online reviews, such as “review manipulation,” “strategic manipulation,” “shill reviews.” Hence, we modified our search string as (“review manipulation” OR “fake reviews” OR “fraud detection” OR “review deception” OR “shill reviews” OR “strategic manipulation” OR “deceptive reviews” OR “sponsored reviews” OR “product-related deception”). Apart from Google Scholar and EBSCO Host, we manually scanned the articles in individual journals such as *MIS Quarterly*, *Information Systems Research*, *Management Science*, *Journal of Management Information Systems*, *Decision Support Systems*, and *Marketing Science* to not miss any important study. Also, to minimize the risk of publication bias, we scanned conference papers on AIS and IEEE conferences. If we found any new paper, we added it to the database.

Next, we assessed the articles manually for quality. We dropped those articles where the search term was either not discussed in detail or was merely a theoretical hook. Notably, we dropped articles discussing deception in general on social media, biases in reviews, herding effect, and financial fraud from further analysis. We now had 80 papers in our basket for review. Finally, we scrutinized the references of the selected papers – papers that did not match our string (such as “collusive reviews,” “fraudulent reviews,” and “opinion spam”). We obtained seven articles through this search. The final basket thus contained 88 articles for our literature review. The domain-wise distribution is as follows: computer science (43.1%), Information Systems (27.27%), Marketing (10.27%), General Management (7.95%), Tourism (6.81%), and Economics and Public Policy (4.45%). Around 21% of these articles were from A* category journals of Australian Business Dean Council (ABDC) journals, namely, *Marketing Science*, *Management Science*, *Decision Support Systems* (DSS), *Journal of Management Information Systems* (JMIS), *The American Economic Review*, *The Economic Journal*, *Information Systems Research* (ISR), *Tourism Management*, *Journal of Travel Research*, *Annals of Tourism Research*. Review manipulation being a relatively new field, has seen the majority of exposure in conferences. 47.7% of articles in our database were from conferences, such as IEEE and AIS. Figure 2 visually depicts the article trend over the 15 years.

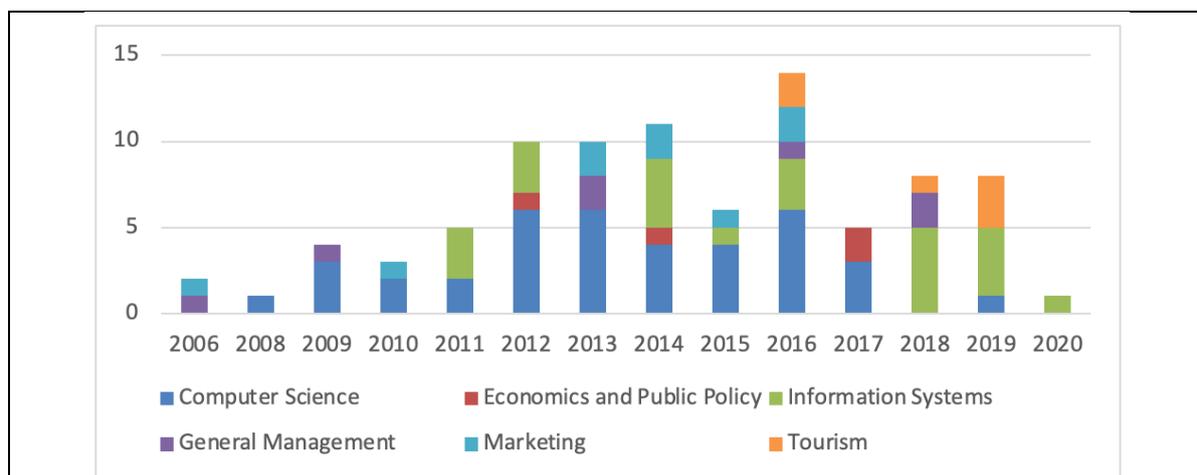


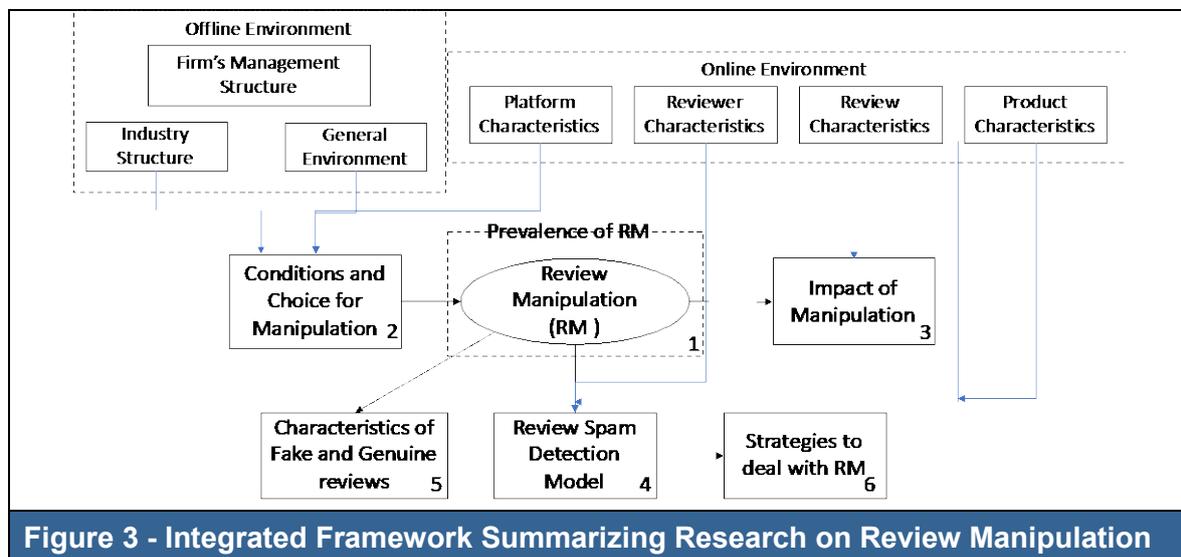
Figure 2 - Articles Published on Review Manipulation Between the Year 2006-2020

We can make the following inferences about noteworthy findings in the literature based on these articles:

- i. Most studies are empirical (e.g., Hu, Bose et al., 2011; Hu et al., 2012; Luca & Zervas, 2016; Mayzlin et al., 2014) and dominated by machine learning approaches (e.g.,

- Jindal & Liu, 2008; Jindal et al., 2010; Mukherjee et al., 2013a; Ott et al., 2012; Ott et al., 2011)
- ii. The literature is mostly atheoretical (e.g., Ananthakrishnan et al., 2020; Dellarocas, 2006; Hu, Bose et al., 2011; Luca & Zervas, 2016; Mayzlin, 2006; Mayzlin et al., 2014)
 - iii. There is a strong focus on firm-level perspective (e.g., Ananthakrishnan et al., 2020; Hu, Bose et al., 2011; Lappas et al., 2016; Mayzlin et al., 2014)
 - iv. The field recognizes the lack of availability of golden-truth data (labeled fake and deceptive reviews) and hence increasingly relies on Yelp.com’s filtering data³ for research.
 - v. The studies primarily cover the hospitality industry (e.g., Lappas et al., 2016; Luca & Zervas, 2016; Mayzlin et al., 2014)

We then coded the articles following an iterative approach (Schryen, 2015) to synthesize the literature on concepts and categorized them independently into potential themes. We then compared the codes and resolved discrepancies through discussions. Through this review, we extracted six prominent themes: Prevalence of manipulation (8.04%), Impact of manipulation (10.34%), Conditions and choice for manipulation decision (5.74%), Characteristics of fake reviews (10.34%), Models for Detecting spam reviews (48.27%), and Strategies to deal with manipulation (17.24%). Some of the articles fell into two or more of these six categories. We categorized such articles based on the primary theme of the paper. However, while discussing the themes in detail in the next section, we utilize knowledge gained from both primary and secondary themes of such articles. By integrating these themes, we also develop an integrated framework that summarizes the work on review manipulation. Figure 3 pictorially classifies the research on review manipulation and presents six major research themes in the area. We derived these themes after multiple iterations and coding. Table 1 summarizes these six themes.



³ Yelp.com is the only platform that publicly displays the reviews detected as fake by its black-box algorithm. Research on review manipulation uses these filtered reviews as dataset for deceptive reviews

Table 1 – Studies Exploring Specific Broadly Defined Themes			
Themes	Key Findings	Illustrative Examples	Theory Used
Box 1: Prevalence of Review Manipulation	Literature finds support that manipulation is a ubiquitous phenomenon on different platforms such as Yelp, Amazon, Expedia, Trip Advisor, Barnes and Noble	Luca & Zervas (2016), Hu et al. (2012); Hu, Liu et al. (2011), Hu, Bose et al. (2011); Kornish (2009); Mayzlin et al. (2014)	Draws parallel from the financial fraud using earnings management literature (Hu, Bose et al., 2011), writing persuading techniques (Hu et al., 2012)
Box 2: Conditions and Choice for Manipulation Decision	Numerous antecedents affect the firm's decision and choice of manipulation, including economic incentives, the online reputation of seller or business, organizational structure, market competition, product quality, price, transaction frequency between seller and buyer.	Mayzlin (2006), Luca & Zervas (2016), Mayzlin et al. (2014), Lappas et al. (2016), Lee et al. (2014), You et al. (2011)	Literature borrows from the advertising domain (Mayzlin et al., 2014); Rational expectation equilibrium (Lee et al., 2014); and literature around visibility construct (Lappas et al., 2016)
Box 3: Impact of manipulation	Manipulation strategy adopted by firms deteriorates forums' informativeness, credibility which eventually leads to drop-in consumer welfare; Manipulation decreases customer's trust; manipulation in the form of volume and valence, and managerial response influences customer's purchase intentions.	Dellarocas (2006), Hu et al. (2012); Hu, Liu et al. (2011), Wang et al. (2016); Xu et al. (2020), Anderson & Magruder (2012), Lappas et al. (2016); Ma et al. (2019)	Reactance theory (Ma et al., 2019), Theory of ethics and impression formation (Hausman et al., 2014), warranting theory (DeAndrea et al., 2018)
Box 4: Review Spam Detection Models	Literature remains inconclusive on the factors that lead to higher accuracy of the models. The research utilizes review-centric features, behavioral and non-verbal features, social features of reviewers along with details on personal information into the model	Kumar, Venugopal, et al. (2018); Kumar, Hooi, et al. (2018), Ott et al. (2012); (Ott et al., 2011), Zhang et al. (2016), Mukherjee et al. (2013b), Rayana & Akoglu (2015), Wang et al. (2011)	Interpersonal Deception Theory (Zhang et al., 2016)
Box 5: Fake and Genuine Review Characteristics	Literature reveals inconsistent patterns while diagnosing genuine and fake reviews. The two vary in terms of readability, subjectivity level, product informativeness, and sentence structure.	Lappas (2012); Martinez-Torres & Toral (2019); Ong et al. (2014); Shan et al. (2018); Yoo & Gretzel (2009)	Deception theory (Yoo & Gretzel, 2009); Ong et al. (2014) used expectancy theory to build the arguments

Table 1 – Studies Exploring Specific Broadly Defined Themes

Box 6: Strategies to deal with manipulation	How does a business deal with manipulation and try to regulate its reputation? Platform-related literature suggests an incentivized reward-based mechanism to foster a fair review submission system.	Ananthakrishnan et al. (2020); Gössling et al. (2018); Gössling et al. (2019); Gutt et al. (2019); Jurca & Faltings (2009); Peng et al. (2016); Ansari & Gupta (2021)	Information Manipulation Theory (Peng et al., 2016)
--	---	---	---

Themes Derived from The Integrated Framework On Review Manipulation

Theme 1: Exploration of Manipulation Prevalence

After several incidences of review manipulation, researchers took the first logical step of exploring the presence of manipulation. They seek an answer to the question – how big is the issue of manipulation? Hu et al. (2012) found that around 10.3% of the book reviews on Amazon and approximately 16% of Yelp reviews are manipulated (Luca & Zervas, 2016). Manipulation on Amazon varies from 20% to 47% (Kornish, 2009). The manipulation is also prevalent on other platforms, such as TripAdvisor, Barnes and Noble, and Expedia. Manipulation on Barnes and Noble is higher than manipulation on Amazon (Hu, Liu et al., 2011).

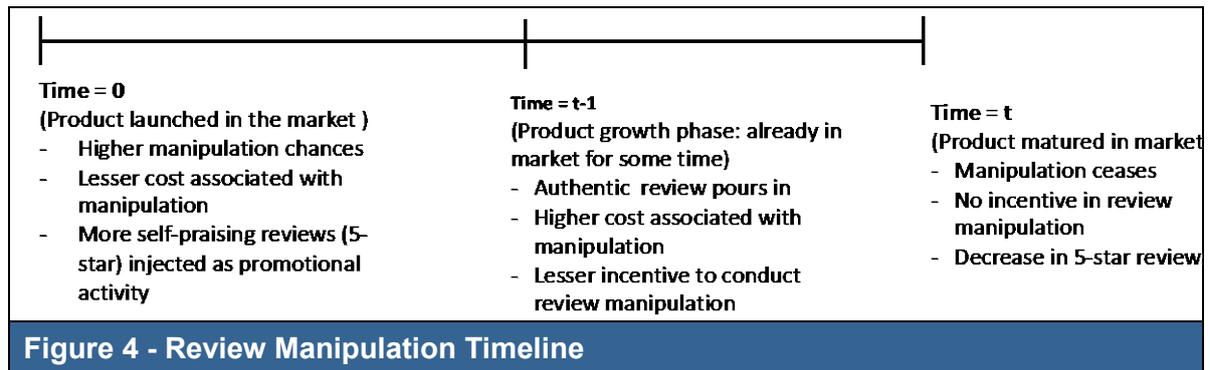
Researchers have developed different proxies to measure manipulation. For instance, Hu et al. (2012) adopted the Wald-Wolfowitz (Runs) test to check the randomness in the writing styles and ratings. Non-randomness signifies manipulation in online reviews. They also considered temporal effects in assessing manipulation. Some studies (e.g., Kornish, 2009; Wan & Nakayama, 2014) propose that the combination of review data and helpfulness count on a review provides a clue to manipulation. Hu, Bose, et al. (2011) proposed a discretionary manipulation proxy to incorporate variability in the ratings. Lesser controllability in the variations caused by manipulation forms the basis of proxy.

The differences-in-differences approach (Mayzlin et al., 2014; Zhuang et al., 2018) and temporal analysis (Hu, Liu et al., 2011) are other techniques for comparing manipulation across websites. Time appears as a crucial factor in manipulation. It emerges from the literature that sellers indulge in manipulation as soon as the product enters the marketplace (Chen & Lin, 2013; Hu, Liu et al., 2011). In due course, manipulation reduces as authentic reviews start pouring in. This is because, with time, writing fake reviews becomes cost-prohibitive for review spammers⁴. Xu et al. (2020) compared netizen reviews⁵ and viewer reviews posted on Naver.com – a South Korean movie database – and found that the positive review ratio's difference immediately increases in the first week of the movie release but then decreases in the following week. They discovered that manipulation of a movie's reviews is likely to occur during the first two weeks of its release. Figure 4 shows the manipulation timeline based on the product lifecycle. We developed it based on the findings from relevant

⁴ Review spammers are those reviewers who indulge in review manipulation through injection of fake reviews.

⁵ Naver.com tags reviews posted by anyone without the purchase of the movie tickets as netizen reviews and the reviews posted by customers of Naver.com are tagged as viewer reviews.

studies. However, it is more suitable for tangible products than intangible ones (such as services). Research on the manipulation cycle for intangible products is still scanty.



Theme 2: Conditions and Choice for Manipulation Decision

The next logical theme in review manipulation that research focuses upon is of identifying the antecedents of manipulation. The questions sought under this theme include: “What makes a firm manipulate? Under what conditions does manipulation occur? What kind of businesses are more susceptible to manipulation or are likely to commit manipulation?” Numerous factors, such as economic incentives, weaker online reputation, organizational structure, competitiveness, and transactional characteristics, drive a seller to commit fraud (Luca & Zervas, 2016; Mayzlin et al., 2014). The research found that online sellers or retailers are more likely to indulge in review manipulation when facing increased competition or when they are not so reputed.

In terms of product category, vendors are more likely to manipulate non-best-selling books, popular books, high-priced books, or books that customers rate as low on the helpfulness scale (Hu, Bose et al., 2011). Li et al. (2017) showed that spammers target mobile apps to increase their average rating. Mayzlin et al. (2014) suggest that restaurants with weak reputations are more likely to commit fraud. Exploring the C2C Chinese market, You et al. (2011) found several factors that promote the likelihood of fake transactions: newly-registered customers to the platform, lesser priced products, sellers with lower reputation but more connectedness with collusion clique, and recurring transaction amongst a seller-buyer pair. They also found that sellers pay pseudo-buyers to conduct fake transactions to become eligible for writing fake reviews. According to Lappas et al. (2016), the self-injection of around fifty fake positive reviews is sufficient to gain visibility over competitors in the market.

Product quality is another crucial factor that affects manipulation. The lower the quality of a product, the higher is the likelihood of manipulation of the reviews. However, manipulation also depends on the cost associated with it. Mayzlin (2006) examined manipulation using game theory and proposed that firms with inferior products find it suboptimal to promote when manipulation costs are high. They should at least perceive some marginal benefit from engaging in review manipulation. Firms would try to manipulate when several users are naïve. The difference between rational and naïve customers is more pronounced when product quality is low (Lee et al., 2014; Lee et al., 2018). In the face of competition, a high-quality firm tends to increase manipulation much more than when its quality is already better than its competitor. On the other hand, a low-quality firm tends to decrease its manipulation level much more when its quality is better than its competitor and vice-versa (Lee et al., 2014; Lee et al., 2018). However, if two sellers sell the same quality products, the one with a lower average rating is more likely to manipulate (Hu, Bose et al., 2011).

A firm’s management structure and geographical location are other indicators of manipulation. According to Mayzlin et al. (2014), independent hotels are more involved in manipulation than

hotels with multi-unit owners or managers. Having another independent hotel nearby increases a hotel's chances of getting injected by negative reviews more than being a multi-unit owner or a branded hotel. Large owner companies or companies with big management are associated with less review manipulation. Further, the platform review verification mechanism sometimes encourages and sometimes discourages promotional reviewing activity (Mayzlin et al., 2014; Xu et al., 2020). Mayzlin et al. (2014) found a significant difference in the review distributions of Trip Advisor (open for everyone to leave a review) and Expedia (allows only those who have booked a hotel and stayed at least for a night to leave a review). Results suggest that Trip Advisor has extreme reviews than Expedia. In short, positive review fraud occurs primarily because of concerns related to reputation, and negative review fraud occurs mostly due to the increased competition (Luca & Zervas, 2016).

Theme 3: Impact of Manipulation

Several studies (e.g., Dellarocas, 2006; Hu, Liu et al., 2011; Lappas et al., 2016) have examined the impact of manipulation on outcome variables such as the platform's informativeness, product sales, customer's welfare, and purchase intentions. Both sellers and platforms have incentives to engage in deceptive reviews as a one full star increase in rating corresponds to more than a 5% increase in sales for online sellers (Luca & Zervas, 2016). Similarly, a half-star increase in rating discourages restaurant's open reservations by 19 percentage points (Anderson & Magruder, 2012). Using an analytical game-theoretic model, Dellarocas (2006) suggests that if a firm inflates its reviews at the rate of its actual quality, manipulation increases the online forums' informativeness. However, if the strategy is a monotonically decreasing function of a firm's actual quality, low-quality firms are more likely to manipulate reviews than high-quality firms. Using this strategy, low-quality firms reduce the gap between their ratings to make it difficult for customers to infer the actual quality. In such settings, manipulation decreases online reviews' informativeness, thus hurting consumers (Dellarocas, 2006; Hu, Liu et al., 2011). Engagement in review manipulation also degrades the information quality and takes a toll on the platform's credibility (Kumar, Venugopal et al., 2018).

Furthermore, in a competitive environment where the quality of a competitor's product is substantially high, the consumer welfare level drops, regardless of the quality of these firms (Lee et al., 2014). But consumer welfare increases significantly if the competitor firm produces somewhat lower quality products. Dellarocas (2006) claimed that customers are smart to adjust their interpretations. Hu, Liu et al. (2011) tested these claims. They found that customers can partially adjust to the manipulation and self-selection bias prevailing in online reviews. Vendors can cheat customers by manipulating the final outcome. Manipulation of sentiments (and not readability and review ratings) increases product sales and the sales rank for a seller on a platform (Hu et al., 2012). This is because it is challenging to decipher the manipulation of sentiment than textual and numeric information. The presence of fake reviews on the platform enhances customers' uncertainty when the percentage of fake reviews on platforms increases or when a customer does not have much prior experience with the platform or reviews (Zhao et al., 2013).

How does manipulation influence a customer's purchase intention? Research suggests that a customers' awareness of the manipulation weakens the effect of manipulation (Bambauer-Sachse & Mangold, 2013; Hausman et al., 2014; Zhuang et al., 2018). Using the theory of ethics and impression formation theory, Hausman et al. (2014) found that customers' purchase intention decreased when they encountered manipulated positive reviews and increased when they discovered manipulated negative reviews. Moreover, when customers find that the reviews are manipulated, their trust in reviews decreases significantly. Zhuang et al. (2018) also found that manipulation by adding positive reviews increased hotel bookings when consumers were unaware of such practices. However, this effect declined when manipulation intensity reached a certain level. After reaching a threshold, customers' suspicion exerted a

negative impact on sales. Bambauer-Sachse & Mangold (2013) showed that knowing about review manipulation weakens the impact of negative reviews, but not when they encounter only positive reviews. Source credibility does moderate the influence of manipulation on purchase intention, such that when a highly credible source provides knowledge, negative reviews do not influence customers' perception.

Next, we discuss the influence of the platform's review verification mechanism on customer's purchase intentions. Xu et al. (2020) found that higher review volume and valence ratio between Expedia and TripAdvisor properties significantly increase traveller bookings in the travel industry. An interesting finding that emerged is that customers treat responses by hotels on TripAdvisor reviews as a signal of manipulation, thus leading to lesser bookings. On the contrary, Ma et al. (2019) found that both higher valence and the greater volume of non-verified reviews for a movie negatively affected its box-office revenue. These contrasting findings suggest the need for re-examination of the role of verification mechanism on customers' decisions.

Theme 4: Review of Spam Detection Models

This is the most established stream of research on review manipulation, which focuses on building data mining algorithms to detect fake reviews. This stream has also received significant interest from computer science scholars. The major challenge in this field is to find a real dataset containing both fake and genuine reviews, though the area has advanced in terms of methodology. Researchers have adopted several approaches to overcome the lack of data. For instance, labeling duplicate reviews (or near duplicates) as deceptive reviews (Jindal & Liu, 2007, 2008); employing Mechanical Turks (Ott et al., 2011) and students (Ong et al., 2014; Yoo & Gretzel, 2009) to write deceptive reviews for an entity; and appointing business students (Wang et al., 2011; Wijnhoven & Pieper, 2018) and expert judges from company officials (Mukherjee et al., 2012) to label reviews as spam, non-spam, and borderline. Several studies (e.g., Kumar, Venugopal et al., 2018; Luca & Zervas, 2016; Mukherjee et al., 2013b; Zhang et al., 2016) have considered filtered reviews from Yelp.com as the dataset. This is because Yelp is the only platform that publicly publishes its filtering algorithm's results on its website. Prior work (Luca & Zervas, 2016; Mukherjee et al., 2013b) has verified that Yelp's filtering algorithm segregates reviews reasonably well. Shojaee et al. (2015) suggest a scientific way to annotate reviews manually. Their method involves taking all the reviews from a particular reviewer at the same time and access each of them based on a clue questionnaire prepared by them for detecting both spam and spammers. Most of the studies have built and tested their deception detection model on reviews from the hospitality industry (Yelp.com, TripAdvisor.com), followed by e-commerce platforms (Amazon.com, consumer electronics businesses listed on Yelp.com).

The literature on spam review detection constitutes two research strands. The first strand of literature focuses on building algorithms to identify fake reviews and spammers and simultaneously look for features that improve detection accuracy (Chen & Lin, 2013; Kumar et al., 2018). The second strand examines if humans can detect fake reviews. These research strands adopt either machine learning (supervised or unsupervised) or non-machine learning techniques (network-based approach or feature engineering). We discuss these two strands of literature in detail.

Building Algorithms for Detecting Fake Reviews

The initial research in this area looked at fake review detection as a classification problem. Scholars used supervised classification models for examining either the review, reviewer (behavioral), product characteristics, or the combination of the two (Chen & Lin, 2013; Zhang et al., 2016). They looked for different patterns, such as duplicate reviews, near-similar reviews (Jindal & Liu, 2008), and burst patterns (Fei et al., 2013; Xie et al., 2012) to identify and label

fake reviews. These studies guide researchers in constructing a training dataset for fake review classifiers.

Jindal and Liu (2008) mark the beginning of fake review detection. They studied duplication of review content to detect spam reviews. They observed that spammers usually create a small number of reviews as templates and duplicate them to spam a single product or different products of the same brand by changing the product name. To build this model, they used review-centric, reviewer-centric, and product-centric features and achieved an accuracy of 78%. Ott et al. (2012) assumed that all first-time reviews are fake. However, questioning this assumption, Xie et al. (2012) suggested looking for bursts in the reviews and proposed a novel way by temporally modeling the reviewer's behavior.

Researchers try out different variations in the existing variables or incorporate new variables to improve the performance accuracy of fake review detection models. Chen and Lin (2013) identified eight attributes that improved the deception detection model's classification accuracy. On raking the correlation coefficient between the review manipulation attributes, sentiment emerged as a key factor in identifying fake reviews. Kumar, Venugopal et al. (2018) incorporated the distributional aspect of review-centric features into various supervised machine learning methods and achieved an AUC score of 0.817 using logistic regression. Review gap distribution contributes the most in detecting spammers. Scholars also suggested other variables, such as user-centric features (Barbado et al., 2019), non-verbal behavior (Zhang et al., 2016), behavioral features (Mukherjee et al., 2013a), psycholinguistic features (Ott et al., 2012), product-related features (Akoglu et al., 2013) and a combination of multiple micro-linguistic cues (Plotkina et al., 2020). Li et al. (2017) and Xu and Zhang (2015) used reviewer posting rates compared with other reviewers to detect collusive fraud. Thus, frameworks have been applied to different languages to identify review spams (Abu Hammad, 2015; Angsumalee et al., 2016; Antonio et al., 2018; Hazim et al., 2018). However, literature still lacks consensus on factors that influence deception and help identify deception in a review.

Building Algorithms for Detecting Review Spammers

Besides identifying fake reviews, research has developed algorithms to identify spammers (those who post spam reviews). Most of the studies on spammer identification (Jindal et al., 2010; Mukherjee et al., 2012; Wang et al., 2011) use the network-centric approach. The network-centric approach refers to the "*analysis of interdependencies through the links or edges between objects (either reviewers or reviews) to obtain the behavior of users in online reviews websites*" (Martinez-Torres & Toral, 2019, p. 394). Akoglu et al. (2013) modeled interactions through a bipartite network which connects reviewer with products via review rating. They extend loopy belief propagation (LBP) for scoring each product to make a ranking list of suspicious users and reviews.

Similarly, Mukherjee et al. (2012) also ranked suspicious spammers and identified a group of spammers using the GSRank algorithm. The interactive computation framework is another algorithm that calculates the trustiness value of reviewers, honesty value of reviews, and reliability value of a product (Wahyuni & Djunaidy, 2016; Wang et al., 2011). Ostensibly, the studies on review detection and deception models strongly assume that future reviews would follow the same pattern. However, if spammers develop their new style to masquerade the reviews, the current detection models would miss them altogether.

Human Detection of Fake Reviews

Prior research (e.g., Levine et al., 1999; Markowitz & Hancock, 2016) on deception argues that humans are poor detectors of deception. They say that humans' deception accuracy may exceed, at most, a little over 50%. Studies on review manipulation have also examined humans' ability to detect fake reviews and compare them with a computer algorithm's accuracy.

Using an experimental design, Plotkina et al. (2020) showed that consumers, even when primed with deception detection tips, fail to distinguish between genuine and fake reviews. The studies in this research highlight how customers miserably fail in detecting fake reviews. Hence, studies can explore technological and non-technological tools to help customers better detect fake reviews.

Theme 5: Fake and Genuine Review Characteristics

Review manipulation, if detected, can cost heavily to the seller as well as the platform. Spammers write reviews in a manner that they do not get noticed. Lappas (2012) suggests that they create fake reviews so that these reviews blend effortlessly with the corpus, have coherence between review rating and text, and are easy to parse. Still, literature has attempted to examine the factors that make a review distinct. Most of the studies (e.g., Luca & Zervas, 2016; Shan et al., 2018) tried to reverse engineer Yelp's filtering algorithm to figure out the characteristics of fake reviews. Broadly, fake reviews tend to be more extreme than others and are written by less reputed reviewers (Luca & Zervas, 2016). Analysis of Yelp's filtered and non-filtered restaurant reviews show inconsistency between product ratings and review sentiment. Findings suggest higher inconsistency (higher sentiment standard deviation) between 1, 4, and 5-star rated fake reviews and genuine reviews. The t-test results also reveal that reviews are inconsistent in terms of length, nouns, verbs, and adjectives. There is a higher positive correlation between sentiment and product rating in genuine reviews (Shan et al., 2018). Ong et al. (2014) observed that fake reviews lack the first-hand experience of the product. They also found that fake reviews do not reveal much about the product and contain more percentages of sentences that meander around the official product features instead of giving their personal opinions (Ong et al., 2014). Consistent with the findings, Martinez-Torres & Toral (2019) found that genuine restaurant reviews focus on the trip's overall experience. The details emphasize other aspects of hotels (such as pet facility, spas, staff behavior, ambiance, cleanliness, etc.) that are not available on the website. Compared to genuine reviews, fake reviews are more lexically complex, include more self-reference, brand name, use more positive words, and lesser negative words (Yoo & Gretzel, 2009). Plotkina et al. (2020) showed that the two review types vary in linguistic structure. They found that genuine reviews are more emotional, easier to read, generally longer (number of paragraphs and sentences per paragraph), more precise in terms of wording. They lack temporal cohesiveness and contain more connectives and fewer adjectives as compared to fake reviews.

Theme 6: Strategies to Deal with Manipulation

This research theme covers studies that discuss customers, platforms, and businesses' responses to review manipulation. Studies from the customer's perspective (e.g., Ansari & Gupta, 2021; Munzel, 2016; Peng et al., 2016) examine how customers respond to different manipulation tactics. Studies from the platform's perspective (e.g., Ananthakrishnan et al., 2020; Edelman, 2017) address this issue in two ways. The first stream explores the tools and platform design that can help customers deal with review manipulation. The second stream (e.g., Jurca & Faltings, 2009; Riazati et al., 2019; Thakur, 2019) explores reward mechanisms that can help foster fair reviews in the review system. Finally, the studies from the businesses' perspective (e.g., Gössling et al., 2019) shed light on how businesses deal with manipulation.

How do customers respond to manipulation? During interviews, customers have revealed that they are aware of manipulation tactics and perceive the addition and deletion of fake reviews as more unethical and deceptive over incentivized tactics for pulling reviews. They rate deletion as the most difficult-to-detect technique (Peng et al., 2016). Customers mostly rely on cues (such as message content and style) and review characteristics (such as review extremity and valence) to check the trustworthiness of reviews (Filiari, 2016). To safeguard customers, Munzel (2016) proposed building deception support mechanisms (such as seals from third-party) for customers. Results show support mechanisms enhance customers' trust

towards reviews and their future intention to return to the website. For example, Yelp has deployed an algorithm that helps detect fraudulent reviews on its website. Ananthakrishnan et al. (2020) takes this forward and seeks to answer the research question: Should platforms display fraudulent reviews? Results show that displaying fraudulent information leads customers to spend more time making their choices and use information rigorously. Display of fraud reviews, along with a trust score, serves as structural assurance and increases consumer's trust on platforms by 80% (Ananthakrishnan et al., 2020).

Apart from providing trust-building mechanisms on platforms, research also proposes building a truthful review system (Gutt et al., 2019; Jurca & Faltings, 2009; Luca, 2017; Riazati et al., 2019; Thakur, 2019). For instance, Thakur (2019) proposes an incentive system to encourage unbiased feedback from buyers and argues to associate a buyer's reputation with sellers. As per this mechanism, a buyer who leaves a review for a seller endorses that particular seller. In case the seller's reputation decrease, the reputation of all the buyers who endorsed this particular seller also decreases. The simulation results suggest that this reputation mechanism encourages rational behavior among buyers and increases the reputation of genuine sellers over non-genuine ones.

Similarly, Riazati et al. (2019) suggest dynamically adjusting the market fee (commission received by the sellers for a transaction) based on the buyer's transaction quality report. A buyer's dishonest reporting will lead to higher remuneration for the seller. Simulation results show that an e-marketplace can become resistant against attacks such as review manipulation and sellers' re-entry through fake-ids by properly implementing reward and punishment mechanisms. Similarly, Ku et al. (2012) suggest a reputation mechanism that platforms can implement to help customers evaluate reputed reviewers. Other strategies suggested by research include revealing reviewers' identity, limiting who can leave reviews, and increasing entry barriers by requiring sign-in, not allowing anonymous posts, and helping customers track a reviewer's activity (Kugler, 2014; Luca, 2017).

Let us now look at the strategies adopted by sellers and platforms to deal with manipulation. Kumar, Qiu et al. (2018) found that the restaurants that respond to customer reviews perform better than others. Responding to the customers could be employed as a strategy by the restaurant managers to safeguard themselves from the consequences of review manipulation. However, responding to reviews has a spillover effect when similar businesses react to the reviews. Managers often are stuck in the rat race, where engaging in manipulation seems the only rational choice in an increasingly competitive market situation (Gössling et al., 2018; Gössling et al., 2019).

Future Research Directions

As discussed above, research on review manipulation is quite fragmented and diverse across various disciplines. In this study, we synthesize them into six different themes. However, the research is still in the nascent stage, and more research is required to gain a comprehensive understanding of review manipulation. Here, we present four suggestions for future research that would strengthen the research in information systems.

Agenda 1: Need for development of a manipulation proxy

As discussed earlier, prior literature has explored manipulation across different platforms, such as Yelp, Expedia, and Amazon; and various industries, such as retailing, hotels, restaurants, and movies. Studies have utilized econometric analysis or temporal-based regression analysis to build a measure of manipulation proxy. Hu, Bose et al. (2011) draws a parallel from the financial fraud literature to develop a manipulation proxy. Scholars have also used differences-in-differences to create their manipulation index. However, with different kinds of forums emerging, research should examine even those forums or platforms that deal

with subjective data over objective truth data (such as forums dealing with political debate, movie reviews, and healthcare platforms). Such platforms can have enormous consequences for customers. Webmed, for example, allows posting reviews of drugs as well as doctors. Manipulated reviews on such platforms can wreak havoc for the public as well as the platform itself. We also need to understand how manipulation works in such an environment. In the case of products on e-commerce or movie business, manipulation ceases after a product reaches a certain time threshold (Ma et al., 2019). The manipulation is at its peak when the product enters the market, and fewer reviews are available for customers to assess it. However, it would be interesting to explore how manipulation would prevail in service provisions (such as hospitals and doctors). Examining this question is essential as it will inform the businesses when they are most susceptible to attack. However, apart from addressing a platform-specific manipulation issue, it is necessary to understand the phenomenon of manipulation in a unified way. This will control for the specific platforms or review medium. Therefore, developing a generic manipulation proxy for quantifying manipulation would help generalize the issue of manipulation.

Agenda 2: Need for focus shift to customers and platforms

We need to shift the focus from the firm's perspective to the customers and the platform's perspective. Prior studies have focused predominately on why, how, who, and when online businesses are likely to commit review fraud. However, they lack informing businesses on safeguarding themselves from review attackers (Lappas et al., 2016). Further, very few studies look at when, why, what, and how customers write and evaluate fake reviews. What leads customers to leave fake reviews? Under what circumstances do they agree to leave customer reviews? Who are these customers? Why do customers leave fake reviews? What motivates them to do so? Some platforms and sellers encourage customers to leave a better review in exchange for an additional product (or a new product sample) or at times offer handwritten notes encouraging people to leave reviews. It would be interesting to see how these ethically accepted practices motivate customers to leave a biased review or fake review. We should note that biased reviews are not fake reviews. We have not discussed online review biases in detail (see Ho et al., 2017 for understanding biases) as it is outside the purview of this study. We also need to focus on the distinction between socially influenced reviews and fake reviews. Is this behavior of leaving fake reviews associated with any particular product category?

Next, we discuss the customer's evaluation of reviews. This is a grey area as the practice of offering products to compensate for reviews is not considered unethical and does not fall under the purview of review manipulation. Research has not explored this area much. Therefore, it would be interesting to examine how customers perceive these reviews written under the umbrella of ethically accepted practices. Some other questions that research can examine are as follows: Do customers discount for review manipulation when evaluating a product? What is the influence of review manipulation on customers' decision-making? Is there a difference in customer's evaluation of actual fraudulent reviews versus perceived fraudulent reviews? These are some of the questions that we can focus upon for future research.

Agenda 3: Need for a theoretical understanding

Despite the widespread need to bring theoretical models and concepts within research, most studies in this field have formulated results using analytical modeling and direct empirical analysis. What is left unexplained in the mechanism of deriving conclusions from the empirical data? Thus, future research may establish this relationship between data and results using a theoretical anchor. Further, as varied platforms and industries suffer from review manipulation, it is necessary to develop a unified theory of manipulation. The theory can abstract away from specific platforms or review medium and instead focuses on fundamental psychological constructs that explain manipulation.

Agenda 4: Need for exploring a policy angle

The menace of fake reviews probably needs regulators or the platform's governance. As technology develops, so is the spammers' ability to generate more fake reviews. As discussed in earlier sections, platforms have built various defense mechanisms such as helpful count, verified badges to curb fake reviews. But these mechanisms have been exploited by spammers and online businesses by bribing customers to rate a review helpful or appointing their friends and relatives to buy a verified reviewer badge. Thus, it becomes necessary for platforms to self-govern themselves and build policies to deal with fake reviews. The current research on review manipulation has offered suggestions on building incentivized review systems or building trusting mechanisms to tackle the review manipulation issue. But it still lacks exploring how platforms' regulation can help curb the issue and increase the trust towards platforms.

Conclusions

This study reviews the current status of the research on review manipulation from a multidisciplinary perspective, focusing on fostering further research in information systems. We present an integrated framework that synthesizes the existing literature spanning 15 years across different disciplines into six themes. These are the *prevalence of review prevalence*, *impact of manipulation*, *conditions and choice for manipulation decision*, *characteristics of a fake review*, *review spam detection model*, and *strategies to deal with manipulation*. The literature reveals that most of the studies in this domain are empirical and focus primarily on econometrics and machine learning techniques.

Overall, this study enhances the understanding of review manipulation by examining what constitutes and characterizes review manipulation and presents its definition. Specifically, we extrapolate the characteristics of deception and present a comprehensive definition of review manipulation. Second, we synthesize the research articles and distill key insights regarding various questions related to review manipulation. An integrated framework synthesizes the literature themes and presents the linkages between multiple variables and themes. Third, we identify future research avenues after synthesizing the literature from a multidisciplinary perspective to foster research in information systems.

Along with contributing to the research, this study also has implications for the practice. This literature review provides holistic information on review manipulation to different stakeholders such as customers, platforms, and sellers. The research highlights the tradeoff for the media between sales and reputation. This study endeavors to inform various stakeholders to deal with the menace of fake reviews to build a healthy and sustainable environment.

References

- Abu Hammad, A. S. (2015). An approach for detecting spam in Arabic opinion reviews. *The International Arab Journal of Information Technology*, 12(1), 9-16.
- Akoglu, L., Chandy, R., & Faloutsos, C. (2013). Opinion fraud detection in online reviews by network effects. In *Proceedings of the International AAAI Conference on Web and Social Media*, 7(1), 2-11.
- Ananthakrishnan, U. M., Li, B., & Smith, M. D. (2020). A Tangled Web: Should Online Review Portals Display Fraudulent Reviews? *Information Systems Research*, 31(3), 950-971.
- * Anderson, E., & Simester, D. (2013). Deceptive Reviews: The Influential Tail. *Working paper, Sloan School of Management*.
- * Anderson, E. T., & Simester, D. I. (2014). Reviews without a purchase: Low ratings, loyal customers, and deception. *Journal of Marketing Research*, 51(3), 249-269.
- Anderson, M., & Magruder, J. (2012). Learning from the crowd: Regression discontinuity estimates of the effects of an online review database. *The Economic Journal*, 122(563), 957-989.
- Angsumalee, B., Soththisopha, N., & Vateekul, P. (2016, October). A framework to detect unqualified restaurant reviews. In *2016 Eighth International Conference on Knowledge and Systems Engineering (KSE)* (pp. 115-120). IEEE.
- Ansari, S., & Gupta, S. (2021). Customer perception of the deceptiveness of online product reviews: A speech act theory perspective. *International Journal of Information Management*, 57, 102286.
- Antonio, N., de Almeida, A. M., Nunes, L., Batista, F., & Ribeiro, R. (2018). Hotel online reviews: creating a multi-source aggregated index. *International Journal of Contemporary Hospitality Management*.
- Bambauer-Sachse, S., & Mangold, S. (2013). Do consumers still believe what is said in online product reviews? A persuasion knowledge approach. *Journal of Retailing and Consumer Services*, 20(4), 373-381.
- Barbado, R., Araque, O., & Iglesias, C. A. (2019). A framework for fake review detection in online consumer electronics retailers. *Information Processing & Management*, 56(4), 1234-1244.
- Chen, L. S., & Lin, J. Y. (2013, July). A study on review manipulation classification using decision tree. In *2013 10th international conference on service systems and service management* (pp. 680-685). IEEE.
- Chen, Y. (2017). Confessions of a paid Amazon review writer. *Digiday*. Retrieved from <https://digiday.com/marketing/vendors-ask-go-around-policy-confessions-top-ranked-amazon-review-writer/>
- Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3), 345-354.
- Chua, A. Y., & Banerjee, S. (2015). Understanding review helpfulness as a function of reviewer reputation, review rating, and review depth. *Journal of the Association for Information Science and Technology*, 66(2), 354-362.
- * Cocarascu, O., & Toni, F. (2016, August). Detecting deceptive reviews using argumentation. In *Proceedings of the 1st International Workshop on AI for Privacy and Security*, 1-8.
- Connors, L., Mudambi, S. M., & Schuff, D. (2011, January). Is it the review or the reviewer? A multi-method approach to determine the antecedents of online review helpfulness. In *2011 44th Hawaii International Conference on System Sciences*, 1-10. IEEE.

- DeAndrea, D. C., Van Der Heide, B., Vendemia, M. A., & Vang, M. H. (2018). How people evaluate online reviews. *Communication Research*, 45(5), 719-736.
- Dellarocas, C. (2006). Strategic manipulation of internet opinion forums: Implications for consumers and firms. *Management Science*, 52(10), 1577-1593.
- Edelman, B. (2017). The market design and policy of online review platforms. *Oxford Review of Economic Policy*, 33(4), 635-649.
- * Elmurugi, E., & Gherbi, A. (2017, August). An empirical study on detecting fake reviews using machine learning techniques. In *2017 7th International Conference on Innovative Computing Technology (INTECH)*, 107-114. IEEE.
- Fei, G., Mukherjee, A., Liu, B., Hsu, M., Castellanos, M., & Ghosh, R. (2013). Exploiting burstiness in reviews for review spammer detection. *Icwsn*, 13, 175-184.
- * Feng, S., Banerjee, R., & Choi, Y. (2012, July). Syntactic stylometry for deception detection. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 171-175.
- * Feng, S., Xing, L., Gogar, A., & Choi, Y. (2012). Distributional Footprints of Deceptive Product Reviews. *Icwsn*, 12(98), 105.
- Ferrara, E. (2019). The history of digital spam. *Communications of the ACM*, 62(8), 82-91.
- Filieri, R. (2016). What makes an online consumer review trustworthy? *Annals of Tourism Research*, 58, 46-64.
- Gössling, S., Hall, C. M., & Andersson, A.-C. (2018). The manager's dilemma: a conceptualization of online review manipulation strategies. *Current Issues in Tourism*, 21(5), 484-503.
- Gössling, S., Zeiss, H., Hall, C. M., Martin-Rios, C., Ram, Y., & Grøtte, I.-P. (2019). A cross-country comparison of accommodation manager perspectives on online review manipulation. *Current Issues in Tourism*, 22(14), 1744-1763.
- Gutt, D., Neumann, J., Zimmermann, S., Kundisch, D., & Chen, J. (2019). Design of review systems—A strategic instrument to shape online reviewing behavior and economic outcomes. *The Journal of Strategic Information Systems*, 28(2), 104-117.
- Hausman, A., Ma, Y. J., & Lee, H.-H. (2014). Consumer responses toward online review manipulation. *Journal of Research in Interactive Marketing*.
- Hazim, M., Anuar, N. B., Ab Razak, M. F., & Abdullah, N. A. (2018). Detecting opinion spams through supervised boosting approach. *PloS one*, 13(6), e0198884.
- Heydari, A., ali Tavakoli, M., Salim, N., & Heydari, Z. (2015). Detection of review spam: A survey. *Expert Systems with Applications*, 42(7), 3634-3642.
- Ho, Y.-C., Wu, J., & Tan, Y. (2017). Disconfirmation effect on online rating behavior: A structural model. *Information Systems Research*, 28(3), 626-642.
- * Hovy, D. (2016, August). The enemy in your own camp: How well can we detect statistically-generated fake reviews—an adversarial study. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 351-356.
- Hu, N., Bose, I., Gao, Y., & Liu, L. (2011). Manipulation in digital word-of-mouth: A reality check for book reviews. *Decision Support Systems*, 50(3), 627-635.
- Hu, N., Bose, I., Koh, N. S., & Liu, L. (2012). Manipulation of online reviews: An analysis of ratings, readability, and sentiments. *Decision Support Systems*, 52(3), 674-684
- Hu, N., Liu, L., & Sambamurthy, V. (2011). Fraud detection in online consumer reviews. *Decision Support Systems*, 50(3), 614-626.

- Jindal, N., & Liu, B. (2007). Review spam detection. In *Proceedings of the 16th International Conference on World Wide Web*, 1189-1190.
- Jindal, N., & Liu, B. (2008). Opinion spam and analysis. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, 219-230.
- Jindal, N., Liu, B., & Lim, E. P. (2010). Finding unusual review patterns using unexpected rules. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, 1549-1552.
- John, C. (2019). Don't Believe Reviews When Shopping Online, Many Are Fake. *The Quint*. Retrieved from <https://www.thequint.com/tech-and-auto/fake-customer-reviews-scam-on-amazon-flipkart>
- Jurca, R., & Faltings, B. (2009). Mechanisms for making crowds truthful. *Journal of Artificial Intelligence Research*, 34(1), 209-253.
- Kelly, T. (2019). Exposed: How rogue firms sell FAKE glowing Amazon reviews to online retailers for £13 each, duping millions of customers into buying shoddy goods. *Daily Mail*. Retrieved from <https://www.dailymail.co.uk/news/article-7819903/Millions-families-duped-buying-shoddy-goods-Amazon.html>
- King, R. A., Racherla, P., & Bush, V. D. (2014). What we know and don't know about online word-of-mouth: A review and synthesis of the literature. *Journal of Interactive marketing*, 28(3), 167-183.
- Kornish, L. J. (2009). Are user reviews systematically manipulated? Evidence from the helpfulness ratings. *Leeds School of Business Working Paper*.
- Ku, Y.-C., Wei, C.-P., & Hsiao, H.-W. (2012). To whom should I listen? Finding reputable reviewers in opinion-sharing communities. *Decision Support Systems*, 53(3), 534-542.
- Kugler, L. (2014). Keeping online reviews honest. *Communications of the ACM*, 57 (11), 20-23.
- Kumar, N., Qiu, L., & Kumar, S. (2018). Exit, voice, and response on digital platforms: An empirical investigation of online management response strategies. *Information Systems Research*, 29(4), 849-870.
- Kumar, N., Venugopal, D., Qiu, L., & Kumar, S. (2018). Detecting review manipulation on online platforms with hierarchical supervised learning. *Journal of Management Information Systems*, 35(1), 350-380.
- Kumar, S., Hooi, B., Makhija, D., Kumar, M., Faloutsos, C., & Subrahmanian, V. (2018). *Rev2: Fraudulent user prediction in rating platforms*. Paper presented at the Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining.
- Kumar, S., & Shah, N. (2018). False information on web and social media: A survey. *Social Media Analytics: Advances and Applications*.
- Lappas, T. (2012, June). Fake reviews: The malicious perspective. In *International Conference on Application of Natural Language to Information Systems*, 23-34. Springer, Berlin, Heidelberg.
- Lappas, T., Sabnis, G., & Valkanas, G. (2016). The impact of fake reviews on online visibility: A vulnerability assessment of the hotel industry. *Information Systems Research*, 27(4), 940-961.
- Lee, S.-Y., Qiu, L., & Whinston, A. (2014). Manipulation: Online platforms' inescapable fate. Paper presented at the *Thirty Fifth International Conference on Information Systems*, Auckland
- Lee, S. Y., Qiu, L., & Whinston, A. (2018). Sentiment manipulation in online platforms: An analysis of movie tweets. *Production and Operations Management*, 27(3), 393-416.

- Levine, T. R., Park, H. S., & McCornack, S. A. (1999). Accuracy in detecting truths and lies: Documenting the “veracity effect”. *Communications Monographs*, 66(2), 125-144.
- Li, S., Caverlee, J., Niu, W., & Kaghazgaran, P. (2017, August). Crowdsourced app review manipulation. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1137-1140.
- * Li, Y., Feng, X., & Zhang, S. (2016, July). Detecting fake reviews utilizing semantic and emotion model. In *2016 3rd international conference on information science and control engineering (ICISCE)*, 317-320. IEEE.
- * Liang, D., Liu, X., & Shen, H. (2014, October). Detecting spam reviewers by combing reviewer feature and relationship. In *Proceedings 2014 International Conference on Informative and Cybernetics for Computational Social Systems (ICCSS)*, 102-107. IEEE..
- * Lim, E. P., Nguyen, V. A., Jindal, N., Liu, B., & Lauw, H. W. (2010, October). Detecting product review spammers using rating behaviors. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, 939-948.
- * Lin, Y., Zhu, T., Wu, H., Zhang, J., Wang, X., & Zhou, A. (2014, August). Towards online anti-opinion spam: Spotting fake reviews from the review sequence. In *2014 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM 2014)*, 261-264. IEEE.
- Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, 29(3), 458-468.
- * Liu, P., Xu, Z., Ai, J., & Wang, F. (2017, July). Identifying indicators of fake reviews based on spammer's behavior features. In *2017 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C)* 396-403. IEEE.
- Luca, M. (2017). Designing online marketplaces: Trust and reputation mechanisms. *Innovation Policy and the Economy*, 17(1), 77-93.
- Luca, M., & Zervas, G. (2016). Fake it till you make it: Reputation, competition, and Yelp review fraud. *Management Science*, 62(12), 3412-3427.
- Ma, H., Kim, J. M., & Lee, E. (2019). Analyzing dynamic review manipulation and its impact on movie box office revenue. *Electronic Commerce Research and Applications*, 35, 100840.
- Markowitz, D. M., & Hancock, J. T. (2016). Linguistic obfuscation in fraudulent science. *Journal of Language and Social Psychology*, 35(4), 435-445.
- Martinez-Torres, M., & Toral, S. (2019). A machine learning approach for the identification of the deceptive reviews in the hospitality sector using unique attributes and sentiment orientation. *Tourism Management*, 75, 393-403.
- Mayzlin, D. (2006). Promotional chat on the Internet. *Marketing Science*, 25(2), 155-163.
- Mayzlin, D., Dover, Y., & Chevalier, J. (2014). Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review*, 104(8), 2421-2455.
- McCornack, S. A. (1992). Information manipulation theory. *Communications Monographs*, 59(1), 1-16.
- Moe, W. W., & Schweidel, D. A. (2012). Online product opinions: Incidence, evaluation, and evolution. *Marketing Science*, 31(3), 372-386.
- Mudambi, S. M., & Schuff, D. (2010). Research note: What makes a helpful online review? A study of customer reviews on Amazon. com. *MIS Quarterly*, 34(1), 185-200.

- Mukherjee, A., Liu, B., & Glance, N. (2012, April). Spotting fake reviewer groups in consumer reviews. In *Proceedings of the 21st international conference on World Wide Web*, 191-200.
- Mukherjee, A., Venkataraman, V., Liu, B., & Glance, N. (2013). Fake review detection: Classification and analysis of real and pseudo reviews. *UIC-CS-03-2013. Technical Report*.
- Mukherjee, A., Venkataraman, V., Liu, B., & Glance, N. (2013b). What yelp fake review filter might be doing? Paper presented at the *Seventh international AAAI conference on weblogs and social media*.
- Munzel, A., & Toulouse, I. A. E. The Malicious Practice of Fake Reviews-Experimental Investigations of Consumers' Deception Detection Support.
- Munzel, A. (2016). Assisting consumers in detecting fake reviews: The role of identity information disclosure and consensus. *Journal of Retailing and Consumer Services*, 32, 96-108.
- Ngai, E. W., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, 50(3), 559-569.
- Nikolic, I. (2020). Amazon deletes 20,000 reviews written by seven of its top 10 reviewers after 'users made thousands of pounds posting fake five star write-ups'. *Daily Mail*. Retrieved from <https://www.dailymail.co.uk/news/article-8701163/Amazon-deletes-20-000-reviews-users-post-fake-five-star-write-ups.html>
- Ong, T., Mannino, M., & Gregg, D. (2014). Linguistic characteristics of shill reviews. *Electronic Commerce Research and Applications*, 13(2), 69-78.
- Ott, M., Cardie, C., & Hancock, J. (2012). Estimating the prevalence of deception in online review communities. Paper presented at the *Proceedings of the 21st international conference on World Wide Web*.
- Ott, M., Choi, Y., Cardie, C., & Hancock, J. T. (2011). Finding deceptive opinion spam by any stretch of the imagination. Paper presented at the *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies*, 1.
- Peng, L., Cui, G., Zhuang, M., & Li, C. (2016). Consumer perceptions of online review deceptions: an empirical study in China. *Journal of Consumer Marketing*, 33(4), 269 - 280.
- Plotkina, D., Munzel, A., & Pallud, J. (2020). Illusions of truth—Experimental insights into human and algorithmic detections of fake online reviews. *Journal of Business Research*, 109, 511-523.
- Rayana, S., & Akoglu, L. (2015). Collective opinion spam detection: Bridging review networks and metadata. Paper presented at the *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- Riazati, M., Shajari, M., & Khorsandi, S. (2019). An incentive mechanism to promote honesty among seller agents in electronic marketplaces. *Electronic Commerce Research*, 19(1), 231-255.
- * Robinson, R., Goh, T.-T., & Zhang, R. (2012). Textual factors in online product reviews: a foundation for a more influential approach to opinion mining. *Electronic Commerce Research*, 12(3), 301-330.
- Schryen, G. (2015). Writing qualitative is literature reviews—guidelines for synthesis, interpretation, and guidance of research. *Communications of the Association for Information Systems*, 37(1), 286-325.

- Shan, G., Zhang, D., Zhou, L., Suo, L., Lim, J., & Shi, C. (2018). Inconsistency Investigation between Online Review Content and Ratings. Paper presented at the *Twenty-fourth Americas Conference on Information Systems*, New Orleans.
- Shojaee, S., Azman, A., Murad, M., Sharef, N., & Sulaiman, N. (2015). A framework for fake review annotation. Paper presented at the *Proceedings of the 2015 17th UKSIM-AMSS International Conference on Modelling and Simulation*.
- Smithers, R. (2019). Flood of 'fake' five-star reviews inundating Amazon, Which? finds. *The Guardian*. Retrieved from <https://www.theguardian.com/technology/2019/apr/16/flood-of-fake-five-star-reviews-inundating-amazon-which-finds>
- Thakur, S. (2019). A reputation management mechanism that incorporates accountability in online ratings. *Electronic Commerce Research*, 19(1), 23-57.
- Wahyuni, E. D., & Djunaidy, A. (2016). Fake review detection from a product review using modified method of iterative computation framework. Paper presented at the *MATEC Web of Conferences*.
- Wan, Y., & Nakayama, M. (2014). The reliability of online review helpfulness. *Journal of Electronic Commerce Research*, 15(3), 179.
- Wang, G., Xie, S., Liu, B., & Philip, S. Y. (2011). Review graph based online store review spammer detection. Paper presented at the *2011 IEEE 11th International Conference on Data Mining*.
- Wang, M., Lu, Q., & Ye, Q. (2016, June). The Impact of Different types of Online Reviews on Consumer Purchasing Decision-an Empirical Investigation with Online Marketplace Data. In *PACIS* (p. 300).
- Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, 26(2), xiii-xxiii.
- Wijnhoven, F., & Pieper, A. T. (2018). Review spam criteria for enhancing a review spam detector. Paper presented at the *Thirty Ninth International Conference on Information Systems*, San Francisco.
- * Wu, G., Greene, D., Smyth, B., & Cunningham, P. (2010). Distortion as a validation criterion in the identification of suspicious reviews. Paper presented at the *Proceedings of the first workshop on social media analytics*.
- Wu, Y., Ngai, E. W., Wu, P., & Wu, C. (2020). Fake online reviews: Literature review, synthesis, and directions for future research. *Decision Support Systems*, 132, 113280.
- Xiao, B., & Benbasat, I. (2011). Product-related deception in e-commerce: a theoretical perspective. *MIS Quarterly*, 35(1), 169-196.
- Xie, S., Wang, G., Lin, S., & Yu, P. S. (2012). Review spam detection via temporal pattern discovery. Paper presented at the *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- Xu, C., & Zhang, J. (2015). Towards collusive fraud detection in online reviews. Paper presented at the *2015 IEEE International Conference on Data Mining*.
- Xu, Y., Zhang, Z., Law, R., & Zhang, Z. (2020). Effects of online reviews and managerial responses from a review manipulation perspective. *Current Issues in Tourism*, 23(17), 2207-2222.
- * Yang, Y., Sun, Y. L., Kay, S., & Yang, Q. (2009). Defending online reputation systems against collaborative unfair raters through signal modeling and trust. Paper presented at the *Proceedings of the 2009 ACM symposium on Applied Computing*.
- Yoo, K. H., & Gretzel, U. (2009). Comparison of deceptive and truthful travel reviews. *Information and Communication Technologies in Tourism*, 37-47.

- You, W., Liu, L., Xia, M., & Lv, C. (2011). Reputation inflation detection in a Chinese C2C market. *Electronic Commerce Research and Applications*, 10(5), 510-519.
- Zhang, D., Zhou, L., Kehoe, J. L., & Kilic, I. Y. (2016). What online reviewer behaviors really matter? Effects of verbal and nonverbal behaviors on detection of fake online reviews. *Journal of Management Information Systems*, 33(2), 456-481.
- Zhao, Y., Yang, S., Narayan, V., & Zhao, Y. (2013). Modeling consumer learning from online product reviews. *Marketing Science*, 32(1), 153-169.
- Zhuang, M., Cui, G., & Peng, L. (2018). Manufactured opinions: The effect of manipulating online product reviews. *Journal of Business Research*, 87, 24-35.

About the Authors

Sana Ansari is a doctoral scholar in the area of Information Systems at the Indian Institute of Management Raipur. Her areas of interest are e-commerce, social media usage for collective action, and online reviews manipulation. Her work has been published in the Pacific Asia Conference on Information Systems (PACIS), and European Conference on Information Systems (ECIS).

Sumeet Gupta is a Professor of Information Systems at Indian Institute of Management Raipur. He received his PhD (Information Systems) as well as MBA from the National University of Singapore. His research interests include electronic governance, e-commerce and business analytics. He has worked on several high profile national and international consultancy assignments such as with SAP A.G., DFS Gallerias, ASEAN secretariat and EDB Singapore. His work has been published in top tier international journals such as Journal of Management Information Systems, Decision Support Systems, European Journal of Operations Research, Omega and at international conferences.