Computers in Human Behavior Reports 10 (2023) 100278

Contents lists available at ScienceDirect

Computers in Human Behavior Reports

journal homepage: www.sciencedirect.com/journal/computers-in-human-behavior-reports

A systematic literature review about the consumers' side of fake review detection – Which cues do consumers use to determine the veracity of online user reviews?

Michelle Walther [a,*], Timo Jakobi [a], Steven James Watson [b], Gunnar Stevens [a]

[a] Bonn-Rhein-Sieg University of Applied Sciences, Institute for Digital Consumption, Department Economic Sciences, Germany
[b] University of Twente, Department of Psychology of Conflict, Risk, and Safety, The Netherlands

ARTICLE INFO

Keywords:
Fake review detection
Fake review cues
User-perspective
Review scam
Opinion scam
Human review fraud detection

ABSTRACT

Background: Consumers rely heavily on online user reviews when shopping online and cybercriminals produce fake reviews to manipulate consumer opinion. Much prior research focuses on the automated detection of these fake reviews, which are far from perfect. Therefore, consumers must be able to detect fake reviews on their own. In this study we survey the research examining how consumers detect fake reviews online.
Methods: We conducted a systematic literature review over the research on fake review detection from the consumer-perspective. We included academic literature giving new empirical data. We provide a narrative synthesis comparing the theories, methods and outcomes used across studies to identify how consumers detect fake reviews online.
Results: We found only 15 articles that met our inclusion criteria. We classify the most often used cues identified into five categories which were (1) review characteristics (2) textual characteristics (3) reviewer characteristics (4) seller characteristics and (5) characteristics of the platform where the review is displayed.
Discussion: We find that theory is applied inconsistently across studies and that cues to deception are often identified in isolation without any unifying theoretical framework. Consequently, we discuss how such a theoretical framework could be developed.

1. Introduction

Private and business life has increasingly migrated to online platforms in recent years (Spithoven, 2020). Likewise, criminals are also increasingly turning to the online world and finding new ways to deceive victims (Van Nek & Bolz, 2021). This includes online shopping fraud (Spithoven, 2020) which can have (severe) physical, emotional and psychological consequences (Button & Cross, 2017; Button, Nicholls, Kerr, & Owen, 2014; Jansen & Leukfeldt, 2018). The emotional and psychological consequences of online shopping fraud include feeling unsafe in the online environment, stressing about finances, and feeling guilty or being blamed for being victimized due to their own behavior (Button & Cross, 2017; Button et al., 2014; Spalek, 2016; Stevens et al., 2023). The experience of being defrauded can lead to feelings of hopelessness, low self-worth or depression (Button & Cross, 2017; Spalek, 2016). The stress can also manifest in (severe) physical consequences such as psychosomatic symptoms or posttraumatic stress disorders which can manifest in skin diseases and insomnia (Button et al., 2014; Spalek, 2016). Due to the increasing use of online shopping, especially since the COVID-19 pandemic (Statista Research Department, 2021), there has also been an increase in online shopping fraud (Chevalier, 2021; Trivedi, 2021). Within online shopping fraud, there are different techniques used by fraudsters to trick consumers, however, we focus on fake user reviews as one of the most prominent examples of consumer fraud (Bambauer-Sachse & Mangold, 2013; Malbon, 2013).

Research has shown that user reviews have a substantial impact on the purchase decision of consumers (Malbon, 2013). Hu, Bose, Koh, and Liu (2012) also show that consumers increasingly rely on these reviews. Therefore, if user reviews on online platforms, such as Amazon, Google, or TripAdvisor, are fake, consumers could be manipulated into buying low quality or unsafe products or services. It is not yet possible to determine the specific costs to consumers of fake review, in part because there are as yet no longitudinal studies indicating the consequences (Spithoven, 2020). However, there are studies showing the costs of

* Corresponding author. Institute for Digital Consumption, University of Applied Sciences Bonn-Rhein-Sieg, Grantham-Allee 20, 53757, Sankt Augustin, Germany.
E-mail address: michelle.walther@h-brs.de (M. Walther).

https://doi.org/10.1016/j.chbr.2023.100278
Received 22 August 2022; Received in revised form 24 February 2023; Accepted 25 February 2023
Available online 27 February 2023
2451-9588/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

shopping fraud more generally. A 2022 representative study by the German federal office for information security shows that eight percent of their participants were victims of online shopping fraud. The consequences of this fraud varied from financial damage, damage to trust in brands and shopping platforms, and time costs (BSI, 2022). Unfortunately, this report could not specify how these costs could be allocated to different cybercrimes, and no differentiation is made between different types of online shopping fraud (BSI, 2022). Nonetheless, given the known high costs associated with shopping fraud generally, the costs of fake online reviews are likely to be substantial (Román, Riquelme, & Iacobucci, 2023).

Fake reviews are reviews that have an intention to mislead customers in their decision to purchase a product. They are often written by reviewers with little or no actual experience with the products or services. Fake reviews can be either unwarrantedly positive, aiming to promote a product, or unjustifiably negative, for example to damage the reputation or sales of competitor products and brands (Zhang et al., 2016). While research on fake review detection has increased in recent years (e.g. Patel & Patel, 2018; Rastogi & Mehrotra, 2017; Santos, Camargo, & Lacerda, 2020), existing literature reviews such as Vidanagama et al. (2020) and Hussain, Turab Mirza, Rasool, Hussain, and Kaleem (2019) have focused on algorithmic, machine learning techniques. Algorithmic fake detection is mostly used by businesses, since the common user does not have the capacity to compare thousands of reviews via an algorithmic approach (Ansari & Gupta, 2021a; Cohen, 2020). Consequently, it is important that we also understand how the typical consumer makes judgements about whether a review is genuine or fake. There are literature reviews that but consider consumers and do not solely focus on algorithmic fake review detection (Ansari & Gupta, 2021a; Wu et al., 2020). Ansari and Gupta (2021a) focus on the umbrella term review manipulation and they review different spam detection models and identify characteristics of fake reviews (e.g. fake reviews lack the first-hand experience of the product). Furthermore, they review the different themes that have been researched in the online review manipulation context. For example, they examine the characteristics and motives of organizations that engage in review manipulation, the impact of review manipulation, and how consumers respond to an online review that they have determined is not genuine. However, they do not consider how consumers detect fake reviews in the first place (Ansari & Gupta, 2021a). Wu et al. (2020) do consider the critical role consumers play in determining how fake reviews impact the market as the ultimate decision makers. However, their review does not consider the decision making processes of consumers. Rather, their review identifies the antecedents (e.g. who writes fake reviews and for what purpose) and consequences (e.g. damaged credibility of reviews and reduced trust in business) of fake reviews, and contrast existing interventions to challenge fake reviews. However, the interventions they review focus on technical and legal interventions such as algorithmic detection of regulatory controls. However, Wu et al. (2020) do conclude that ultimately it is important on which website the review is displayed and if consumers perceive to have an influence on how much fake reviews impact the market.

Therefore, there remains a need to consider ways of informing and protecting consumers from fake reviews (Ansari & Gupta, 2021a). An important step toward supporting consumers is not only to know how they respond to fake reviews (as e.g. Ansari & Gupta, 2021a), but to also know how consumers process cues from online reviews to estimate veracity. This literature review is a first step in gathering the existing literature on how consumers determine if an online review is genuine or not. In particular, we, seek to determine which cues to veracity actually inform consumer decisions, and how those cues are processed by consumers to form a veracity judgment based on the existing literature.

Understanding the processes by which consumers make judgments will allow for the development of more directly tailored interventions. Thus far, there are few websites that help consumers to analyze the credibility of reviews, for example ReviewMeta,[1] Fakespot,[2] The Review Index[3] and Review Sceptic[4] (Cohen, 2020). However, it is unclear what proportion of users are aware of or engage with such tools or what level of protection they confer. In light of the limitations and inaccessibility of technical support systems, it remains important for consumers to be able to distinguish between fake and genuine reviews themselves. However, at present there is no clear overview of how people distinguish fake from genuine reviews. In order to be successful from a socio-technical perspective, websites or algorithms supporting consumers in the veracity judgment of a review must facilitate users' ability to accurately identify relevant detection cues. That is why in this paper we consolidate existing work about fake review detection and in order to identify the detection cues consumers use that have already been identified by prior research.

1.1. State of the art

Researchers commonly distinguish between three different types of fake reviews (Ansari, Gupta, & Dewangan, 2018; Hu et al., 2012; Jindal & Liu, 2008; Mayzlin, Dover, & Chevalier, 2014; Ren & Ji, 2019). The first type are untruthful opinions, which are reviews that present false opinions of products (Jindal & Liu, 2008). The second type are reviews that focus on the brand of the product and not the product itself, which means that they do not actually provide consumers with advice about a specific product (Jindal & Liu, 2008). The last type are non-reviews. These cover advertisements for different products or services and other irrelevant information without containing any opinions (Jindal & Liu, 2008). The second and the third type of fake reviews are easily detected by machines and people, yet most consumers have no idea or are unsure about how to detect the first type of untruthful opinion reviews (Cohen, 2020). In addition, even if people are confident in detecting fake reviews, research shows that humans are only about 60%–80% accurate in labeling user reviews as fake or genuine (Plotkina, Munzel, & Pallud, 2020; Shukla et al., 2019). Compared to machine learning approaches, which can be up to 90% accurate, human judges are far less accurate in detecting fake reviews (Harris, 2012; Hovy, 2016; Kim, Kang, Shin, & Myaeng, 2021; Kronrod, Lee, & Gordeliy, 2017; Masip, Bethencourt, Lucas, Segundo, & Herrero, 2012; Ott, Choi, Cardie, & Hancock, 2011).

Due to the superiority of machine learning over human detection, existing research on fake review detection mostly focuses on automated fake review detection (Ansari et al., 2018; Ezhilarasan, Govindasamy, Akila, & Vadivelan, 2019; Hossain et al., 2020; Hussain et al., 2019; Patel & Patel, 2018; Vidanagama et al., 2020). Since most of the research is on automated fake review detection, and because the cues used to train machine learning approaches to fake review detection may have relevance to how humans detect fake reviews, we will also consider the features that are used to train machine learning approaches to identify fake reviews.

Most features that are used to identify fake reviews can be categorized as (1) Review centric, (2) Reviewer centric, (3) Meta-data centric or (4) Product centric (Heydari, Tavakoli, & Salim, 2015, 2016; Mohawesh et al., 2021; Rastogi, 2020; Rastogi & Mehrotra, 2017; Ren & Ji, 2019). The features from the different categories are sometimes used in combination for better detection accuracy (Ott et al., 2011). Appendix C provides a detailed overview of the features used in existing literature, but we briefly describe each category here.

Review centric features are the most popular feature category used to

[1] https://reviewmeta.com/.
[2] https://www.fakespot.com/.
[3] https://thereviewindex.com/us.
[4] http://reviewskeptic.com/.

train automated fake review detection tools (Alsubari, Shelke, & Deshmukh, 2020, pp. 3846–3856; Budhi, Chiong, & Wang, 2021; Jindal & Liu, 2008; Ott et al., 2011; Reddy, Babu, Scholar, & Tech, 2020). Review centric features are aspects of the reviews themselves that can be indicative of whether a review is genuine or fake. For example, the use of adverbs can indicate whether a review is genuine or fake, because fake reviews generally contain more adverbs than genuine reviews (Ahmed, Traore, & Saad, 2018; Ott et al., 2011; Shukla et al., 2019),.

The second most popular category of feature is the reviewer centric type (Barbado, Araque, & Iglesias, 2019; Fazzolari, Buccafurri, Lax, & Petrocchi, 2020; Fontanarava, Pasi, & Viviani, 2017; Gera, Thakur, & Singh, 2015; Goswami, Park, & Song, 2017; Harris, 2018; Jindal & Liu, 2008; Y. Li, Feng, & Zhang, 2016; Pillala, Bhansali, Reddy, & Rojamani, 2020; Pokharkar, Shete, & Ghogare, 2017, p. 7; Shehnepoor, Salehi, Farahbakhsh, & Crespi, 2017; Wang et al., 2016; Z. Wu et al., 2017). These features assess aspects of the user that posted a review, rather than the content of the review itself. Reviewer centric features include trust features (Barbado et al., 2019; Fazzolari et al., 2020; Goswami et al., 2017; Jindal & Liu, 2008). An example of trust features are personal features of the reviewer (Barbado et al., 2019), such as whether they use their real name (Barbado et al., 2019; Pillala et al., 2020). Recent research on reviewer centric features also focuses on the detection of reviews that were written by the same reviewer under different names (Le, Li, & Li, 2022). This means that some reviewers post the same review under a product, but use different usernames to pretend to be multiple people expressing the same opinion of that product.

The third category are meta-data features which include temporal features (Fontanarava et al., 2017; Khurshid, Zhu, Xu, Ahmad, & Ahmad, 2019), such as the date variance of the reviews (the amount of time that passes between reviews) (Fontanarava et al., 2017).

The fourth category is product centric features (Jindal & Liu, 2008; Wang et al., 2016) for example the price (Jindal & Liu, 2008) or the average rating of the product (Jindal & Liu, 2008; Wang et al., 2016).

Other authors, such as Wu et al. (Z. Wu et al., 2017), include topological (interaction features on social media) and group features in addition to the four most common categories. Furthermore, there are algorithmic approaches that focus on deep learning which do not use specific review-, behavior-, or product centric features (Hernandez, Rahman, Recabarren, & Carbunar, 2018; Li, et al, 2016; Saini & Sharan, 2017; Wu et al., 2020).

This prior work helpfully identifies cues to online review veracity, and provides a classification system for those cues. In this way they provide a useful starting point for the identification and classification of cues used by consumers to detect fake online reviews. However, humans do not process information in an algorithmic way (Korteling et al., 2021) and therefore, there is a need to know if humans use the cues that have been identified in algorithmic approaches, and if they can be categorized similarly. Moreover, identifying and classifying cues would not directly inform us about how consumers process cues to make veracity decisions. It is also likely to remain the case that the ultimate decisions about whether to trust a review or not will lie with the consumer. Algorithmic fake review detection is unlikely to perfectly remove all fake reviews, and automated fake review detection is not realistically accessible for all consumers (Ahmed et al., 2018; Jindal & Liu, 2008; Kennedy, Walsh, Sloka, Foster, & McCarren, 2019). Therefore, there is a clear need to better understand how consumers identify and process cues to deception in online reviews.

1.2. Research question

In light of the shortcomings of machine learning approaches and the high likelihood consumers will encounter fake reviews, supporting consumers in their judgement remains an important task. However, it is not well understood how consumers approach detecting fake reviews in the first place, or how far consumer judgements make use of known cues of deception used within automated detection. In order to develop successful interventions for consumers we first need to understand how they make their own judgements regarding review veracity. Therefore, our research question is:

What **cues and features** are being used by **consumers** for the **detection** of **fake reviews** in online shopping **recommendation systems**?

We address this question via a systematic review of the extant peer reviewed literature and use our findings to discuss a research agenda to further shape the direction of this young research field.

2. Methods

To address our research question we use a systematic review methodology following the guidelines of Kitchenham (2004).

2.1. Inclusion and exclusion criteria

We first determined the scope of our study through our inclusion and exclusion criteria. We present these criteria in Table 1. The inclusion criteria required that the research must focus on fake review detection, be user-centered, and identify cues for detecting online shopping fraud. Therefore, studies which compare the accuracy and performance between automated fake review detection approaches and human judges, without indicating the cues used by the human judges to identify the veracity of a review, were not included in the scope of this study (e.g. Kim et al., 2021; Ott et al., 2011; Plotkina et al., 2020; Yuan et al., 2016). We included papers about any kind of online shopping, for example booking hotels, shopping for clothing, gifts or household goods, or shopping via Marketplace platforms. Included research had to provide new empirical data which had not been previously reported, but we do not limit studies by design. That is, both qualitative and quantitative studies of all types could be included so long as they met our other inclusion criteria. Our exclusion criteria were that we excluded papers which are not peer-reviewed academic literature. This means that citations, patents, books, PhD theses, Master and Bachelor theses were excluded from the literature search. We also excluded everything that is published before 2007, because Jindal & Liu (2007) were the first to provoke wider analysis of fake reviews and their detection. Furthermore, we excluded research that does not provide new empirical data on fake review detection, for example literature reviews. Research about spammer and spammer group identification as well as fake news detection were excluded since they did not fall into the scope of this study. Finally, for pragmatic reasons we excluded papers that were not written in Dutch, English or German, since we were not able to translate other languages into our analysis.

2.2. Data sources and search strategy

The key objective of this research is to identify all studies that focus

Table 1
Inclusion and exclusion criteria for the papers.

Inclusion criteria	Fake review detection
	Quantitative studies when cues are mentioned for detecting fake online reviews
	Qualitative studies when cues are mentioned for detecting fake online reviews
	User-centered
	Must include new empirical data on that identifies human rated cues for identifying fake online reviews
Exclusion criteria	Machine learning where algorithms were used
	Not peer-reviewed academic literature such as books, Master and Bachelor Thesis etc.
	Papers written in a language other than German, Dutch or English
	Citations & patents
	Books
	Articles that could not be accessed despite our best efforts

on how consumers detect fake reviews in online shopping. Before conducting the systematic search, we ran some preliminary searches to identify keywords and synonyms associated with our topic (see Appendix A and see Table 2).

From these synonyms we constructed a search string that included all known synonyms of fake reviews, detection cues and used Boolean operators to distinguish between synonyms and essential keywords. We included the user-perspective because we are focusing on consumers detection techniques and included synonyms for e-commerce. Since we did not want to include automated fake review detection, we excluded those. Our final search string was then entered into the search engines Web of Science, IEEE Xplore and Elsevier ScienceDirect:

> (deception cues OR detection strategies OR deception detection OR spam indicators OR opinion fraud detection OR opinion spam detection OR opinion spam features) AND (opinion scams OR opinion fraud OR fake reviews OR opinion spam OR online review OR review spam) AND (user-perspective OR consumer-perspective OR consumer OR human) AND (recommender systems OR persuasive recommendations OR e-commerce) NOT (machine learning OR autoencoder OR deep learning)

However, Web of Science and IEEE Xplore only produced six results, while Elsevier did not offer any results. Therefore, we decided to broaden the search and use Google Scholar. Here the search string was too long, and we therefore used Harzing's Publish and Perish software[5] to extract the literature. We combined the results from all search engines for our analyses.

2.3. Study selection process

The selection process of this literature review had several steps. In a first step the first author read the titles of the papers and decided whether the title of the paper indicated that the paper might meet the inclusion criteria. Articles were only excluded when the title indicated they were clearly irrelevant. The abstract of the retained articles were then read. Again, articles were excluded at this stage only if it was clear the paper did not fall into the scope of the review, all remaining abstracts were retained for full text screening. Papers were retained at full text reading only where they met all inclusion criteria and no exclusions criteria. The data collection and extraction were performed by only the first author. However, we agreed that if the first author was unsure about whether the study fitted the scope of the review, the other authors would look at the paper and a decision would be made through discussion. However, during the process it was never necessary to enact this since the decisions were clear.

Table 2
Synonyms for the keywords of the research question.

Key words from RQ	Detection Cues	Fake reviews	Recommendation systems	User-perspective
Synonyms	detection strategies deception detection spam indicators Opinion fraud detection opinion spam detection opinion spam features	opinion scams opinion fraud Opinion spam Online review Review spam	persuasive recommendations e-commerce	Consumer-perspective Consumer Human

2.4. Data extraction and data items

The data extraction was conducted in a standardized way (Biolchini, Mian, Natali, & Travassos, 2005; Castelli, Stevens, & Jakobi, 2019). We collected general information such as publication year and country in which the study was done. Additionally, we collected information about the methodology and the theories (if any) that were used. Lastly, we collected the deception cues that were used or identified (depending on the methodology), and the main findings of each study. We used the classification of cues to review veracity identified from algorithmic approaches to fake review detection as our starting point. We present these classifications in Appendix C. Where those categorizations did not apply, we used the descriptions of cues and any categorical classification of those cues provided by the authors of the reviewed papers to determine classification. Where these descriptions did not fit within any of the preexisting categories provided in Appendix C, we clustered cues based on core common characteristics. For example Ansari et al. found that the more influence the seller has on the website the less trustworthy the reviews are perceived to be (2018); since this feature does not fit in the categories of meta-data, reviewer-centric, review-centric or product-centric we grouped it together with the seller reputation and the trade record under the category "seller characteristics". We extracted the data by coding them with the software MAXQDA,[6] allowing us to keep track of and classify all cues to deception identified across all studies in our review. We also used MAXQDA to track the methods and theories used in each reviewed study. We did this by forming a descriptive table in MAXQDA.

2.5. Study risk of bias

Systematic reviews often assess the extent to which individual study findings may be at risk of being biased. However, we did not assess quality with the risk of bias because of the variety of methodologies in the studies which would mean there would be no clear or fair way to compare risk of bias across studies (Higgins et al., 2021). Instead, we limited our analysis to assessing the quality of methodology used across the studies we identified as a whole.

2.6. Synthesis

Our aims were to survey the theories, methods and outcomes used in studies on fake review detection thus far, and to synthesize their commonalities and differences. Therefore, we employed a narrative rather than statistical synthesis. Since we did not perform any meta-analysis due to the variety of different methodologies used in the different studies, we did not ask the authors for original data.

3. Results

3.1. Study selection

Our search strategy initially identified 989 articles. However, in our preliminary search for key words we found 26 articles that we did not find in the systematic search, but their title implied that they might be

[5] Harzing, A.W. (2007) Publish or Perish, available from https://harzing.com/resources/publish-or-perish; Version: 8.1.3683 (6 December 2021). Harzing's Publish and Perish is a software program which allows you to get a clear list of citations that meet the specific search criteria for a specific database. We used the software to retrieve our Google Scholar search and this made it easier to screen articles for relevance from their title and abstract.

[6] Hersteller VERBI GmbH, MAXQDA Letzte Version: Release 2022.0.0, https://www.maxqda.de/download.

relevant for our study. Therefore, we decided to add these articles to those we identified through our keyword search. In total, we identified 1,015 potentially relevant articles for our analysis.

We did four rounds of selection as outlined in Fig. 1. In the first round we scanned the titles of all articles; literature that obviously did not cover fake review detection, duplicates and literature about reviewer detection were excluded in this round (n = 697), which left 318 articles. In the next round of exclusion, the abstracts were examined. We then excluded research about fake news and research focusing on the perpetrators of fake reviews because they also do not fall in the scope of our review (n = 223). Of the 95 articles that remained after abstract screening four were not available online, therefore we contacted the authors. For one paper, the author did not respond to these requests and so Gupta & George (2013) could not be accessed and was excluded. We then examined the full text of 94 articles and after applying the inclusion and exclusion criteria we excluded 79 more papers.

This means that the analysis is based on the following 15 articles, of which five were from the Google Scholar orientation search: Ananthakrishnan, Li, and Smith (2020); Ansari et al. (2018); Ansari and Gupta (2019, 2021b); DeAndrea, Van Der Heide, Vendemia, and Vang (2018); Filieri (2016); Jensen, Averbeck, Zhang, and Wright (2013); Kronrod et al. (2017); Kusumasondjaja, Shanka, and Marchegiani (2012); Luo, Luo, Schatzberg, and Sia (2013); Munzel (2015, 2016); Peng, Cui, Zhuang, and Li (2016); Racherla, Mandviwalla, and Connolly (2012); Román, P. Riquelme, and Iacobucci (2019).

3.2. Study characteristics

The 15 papers we found during our literature search can be categorized into three groups. First we have the papers that directly address our research question by identifying or testing cues consumers use when determining the veracity of an online review (Ansari et al., 2018; Ansari & Gupta, 2019, 2021b; DeAndrea et al., 2018; Kronrod et al., 2017; Peng et al., 2016; Román et al., 2019, pp. 141–166). Secondly, we have papers that did not focus on actual deception cues but on the broader concept of trust (Ananthakrishnan et al., 2020; Filieri, 2016; Munzel, 2015, 2016; Racherla et al., 2012). And lastly, we have papers that did not focus on actual deception cues but on the broader concept of credibility (Jensen et al., 2013; Kusumasondjaja et al., 2012; Luo et al., 2013). For an overview of all papers with the theories and methods they used and the cues they found see Appendix B.

In the following we will discuss the cues found and methods and theories used across the studies.

3.3. Detection cues found to be used by consumers

We found seven papers that concentrated on the identification of fake review detection cues used by consumers (Ansari et al., 2018; Ansari & Gupta, 2019, 2021b; DeAndrea et al., 2018; Kronrod et al., 2017; Peng et al., 2016; Román et al., 2019, pp. 141–166). However, three papers from Ansari and colleagues report on different aspects of the same mixed-methods study. The three reports use different theoretical frameworks and report different aspects of their results. Therefore, we will report the results and theoretical framework separate from each other but will treat the methods as being from a single study.

We found five studies (Ananthakrishnan et al., 2020; Filieri, 2016; Munzel, 2015, 2016; Racherla et al., 2012) that focus on the factors that influence perceived trustworthiness, and three studies (Jensen et al., 2013; Kusumasondjaja et al., 2012; Luo et al., 2013) that focus on the cues that influence the perceived credibility of a reviews. We find that deception cues and cues to trustworthiness and credibility are inseparable from each other. While deception cues give reasons why reviews might be fake, trustworthiness and credibility cues give reasons why reviews might be genuine. In turn, for example, a lack of deception cues could also influence trustworthiness and credibility. Therefore, we considered them all together.

A variety of cues were used, tested or found across all studies (see Table 3 for an overview). However, most cues can be categorized as addressing review characteristics (Ansari et al., 2018; Ansari & Gupta, 2019, 2021b; Filieri, 2016; Jensen et al., 2013; Kronrod et al., 2017; Peng et al., 2016; Racherla et al., 2012; Román et al., 2019, pp. 141–166), textual characteristics (Ansari et al., 2018; Ansari & Gupta, 2021b; Jensen et al., 2013; Kronrod et al., 2017; Román et al., 2019, pp. 141–166), reviewer characteristics (DeAndrea et al., 2018; Kusumasondjaja et al., 2012; Munzel, 2016; Peng et al., 2016), seller characteristics (DeAndrea et al., 2018; Peng et al., 2016; Racherla et al., 2012), or characteristics of the platform where the review is displayed (Ananthakrishnan et al., 2020; Munzel, 2015, 2016). This contrasts with our initial framework of (1) Review centric, (2) Reviewer centric, (3) Meta-data centric and (4) Product centric features derived from the automated review detection literature. This highlights that the ways in which humans determine credibility are meaningfully different to machine learning approaches with only review and reviewer characteristics carrying over into our final taxonomy. Cues that were used or identified across multiple studies are now reviewed per category.

3.3.1. Review characteristics

Review characteristics include the review content, that is, what the review actually says and details about the product, usefulness count of the review (Ansari et al., 2018), and other characteristics of the review itself.

Review extremity, for example the use of 1 or 5 star ratings, was found in four different studies to be an indicator of untruthfulness (Ansari et al., 2018; Ansari & Gupta, 2019, 2021b; Filieri, 2016; Kronrod

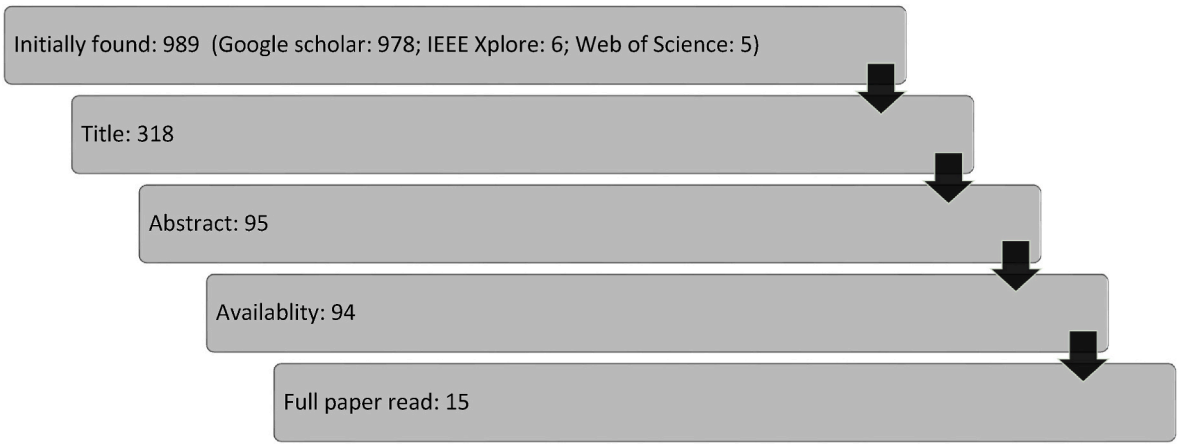

Fig. 1. Number of articles included after each exclusion round.

Table 3
Overview over studies including the cues they found/tested, sorted by year.

Study	Cues	Cue classification	Additional information
Racherla et al. (2012)	• Argument quality • Perceived similarity between subject and reviewer	• Review characteristics • Reviewer characteristics	
Kusumasondjaja et al. (2012)	• Message valence • Reviewer's identity	• Review characteristics • Reviewer characteristics	
Jensen et al. (2013)	• Lexical complexity • Two-sidedness • Affect intensity	• Review characteristics • Review characteristics • Review characteristics	
Luo et al. (2013)	• Recommendation persuasiveness • Recommendation completeness • Recommendation credibility	• Review characteristics • Review characteristics • Review characteristics	
Munzel (2015)	• Displaying of fake reviews • Consensus among reviews • Explanation about detection support mechanism	• Platform characteristics • Review characteristics • Platform characteristics	Persuasion knowledge of the participant was measured as a moderating variable
Filieri (2016)	• Content and writing style of a review message • Valence and review extremity • Source of communication • Pattern in reviews • Receivers' involvement • Consumer experience • Medium • Consumer decision making	• Textual characteristics • Review characteristics • Platform characteristics • Review characteristics • Platform characteristics • Platform characteristics • Platform characteristics • Platform characteristics	
Munzel (2016)	• Identity-descriptive information • Consensus	• Reviewer characteristics • Review characteristics	Persuasion knowledge of the participant was measured as a moderating variable
Peng et al. (2016)	• Quality of content • Quantity of the reviews • Mismatch between review • Seller reputation • Trade record of the seller • Identities of the reviewers	• Review characteristics • Review characteristics • Review characteristics • Seller characteristics • Seller characteristics • Reviewer characteristics	
Kronrod et al. (2017)	• Avoidance of detail • Specific descriptions or names, • Vagueness and omission of information • Overly short or long reviews	• Textual characteristics • Textual characteristics • Textual characteristics • Review characteristics • Review characteristics	Participants were given cues for help in some conditions, the study found that giving cues to participants makes them

Table 3 (*continued*)

Study	Cues	Cue classification	Additional information
	• Extreme valence (overly, negative/positive/good/bad) • Lack of descriptions of personal experience • Presence of grammar/spelling errors	• Reviewer characteristics • Textual characteristics	overly suspicious
Ansari et al. (2018)	• Review extremity • Helpfulness • Informativeness of the website • Subjectivity and Objectivity • Writing style (use of pronouns, adverbs) • Amount of detail • Spelling of review • Grammar of review	• Review characteristics • Review characteristics • Platform characteristics • Textual characteristics • Textual characteristics • Textual characteristics • Textual characteristics • Textual characteristics • Textual characteristics	The dependent variable was the impression of a review
DeAndrea et al. (2018)	○ Platform where the review appears (review website such as Yelp.com or company website)	• Platform characteristics	Knowledge about the general existing of review spamming has an impact on the cue attention (priming) In the other condition, the same reviews, originating from Yelp.com, appeared on the restaurant's own website.
Ansari and Gupta (2019)	Central cues • Argument quality • Source quality • Message length Peripheral cues • Verified badge • Most recent reviews • 1& 5 star ratings	• Review characteristics • Review characteristics • Review characteristics • Review characteristics • Review characteristics • Review characteristics	
Román et al. (2019)	Central cues • Argument quality • Review quantity Peripheral cues • Review quantity • Review consistency • Review homophily,	• Review characteristics • Review characteristics • Review characteristics • Review characteristics • Reviewer characteristics	Also moderator variables were taken into account, such as purchase context, the review platform, hedonic versus utilitarian purchase and purchase involving negative needs versus positive want

(*continued on next page*)

Table 3 (*continued*)

Study	Cues	Cue classification	Additional information
Ananthakrishnan et al. (2020)	• Summarized version of the fraud review information in the form of a trust score	• Platform characteristics	Also it was tested what influence the display of Fake review has on the participants trust in the product
Ansari and Gupta (2021b)	• Referencing • Flattering • Contextual embedding • Detailing • Argument structure	• Textual characteristics • Textual characteristics • Textual characteristics • Textual characteristics • Review characteristics	

et al., 2017). Argument quality (persuasiveness) was found by four studies (Ansari & Gupta, 2019; Luo et al., 2013; Racherla et al., 2012; Román et al., 2019, pp. 141–166) to be an indicator of genuine reviews, with higher quality arguments indicating genuineness. Ansari et al. (Ansari et al., 2018; Ansari & Gupta, 2019) and Kronrod et al. (2017) both found that if a review is overly short or long, it is perceived as more suspicious. Filieri (2016) and Kusumasondjaja et al. (2012) found that the valence of a review is by some seen as a cue to detect the veracity of a review, meaning that negative online reviews are perceived as more credible than positive ones. However, a positive review has more influence on initial trust than a negative one. Kusumasondjaja et al. (2012). Jensen et al. (2013), on the other hand, found that two-sided reviews are more credible. They furthermore found that the complexity of language does not influence the credibility assessment of the review while a high affective intensity decreases review credibility (Jensen et al., 2013). Ansari et al. (2018) found that a higher review helpfulness count for a review is more likely to be perceived as genuine. Additionally, Ansari and Gupta (2019) find peripheral cues, such as verified badges, the most recent reviews and 1 and 5 star ratings, which influence the assessment of a review towards the positive (verified badges and the most recent reviews) and negative (1 and 5 star ratings).

Multiple studies show that reviews which correspond with other reviews, in content and rating, are perceived to be more genuine (Filieri, 2016; Munzel, 2015, 2016; Peng et al., 2016; Román et al., 2019, pp. 141–166). In this vein, Filieri (2016) and Munzel (2016) show that the number of reviewers complementing or complaining about the same things positively influences the reviews' perceived plausibility (Filieri, 2016; Munzel, 2016). Essentially, consistent reviews provide social proof, which is known to impact on purchasing decisions (Amblee & Bui, 2011). However, if the wording of the reviews is too similar, consumers perceive them as fabricated and therefore less genuine (Peng et al., 2016). Furthermore, Peng et al. (2016) and Román et al. (2019, pp. 141–166) found that the quantity of the reviews per product has a positive influence of the perceived veracity of a review.

Román et al. (2019) and Peng et al. (2016) found that the review quantity has a positive influence on the perceived trustworthiness of the product, which means that consumers perceive a product as more trustworthy the more reviews the product has.

3.3.2. Textual characteristics

Textual characteristics include writing style, such as the use of pronouns and adverbs, but also the amount of detail and the spelling and grammar of the review (Ansari et al., 2018; Filieri, 2016). Research found that textual characteristics can influence the perceived veracity of a review (Ansari et al., 2018; Ansari & Gupta, 2021b; Kronrod et al., 2017).

Kronrod et al. (2017) found that people assess a review based on avoidance of detail, specific descriptions or names, vagueness, omission of information, and the presence of grammar/spelling errors (Kronrod et al., 2017).

Ansari and Gupta (2021b) find that referencing, flattering, contextual embedding, and detailing influence the perceived veracity of a review. The authors define "referencing" as providing cues which identify the people involved, for example, the use of first person pronouns. Ansari and Gupta (2021b) find that reviews containing more personal pronouns are perceived as less genuine. The authors argue that such a counter-intuitive outcome could be due to a relational connection that the reader makes between the reviewer and the product, which enhances the subjectivity and therefore the believability (Ansari & Gupta, 2021b). In a different study Ansari et al. (2018) also find that reviews with a higher level of subjectivity are perceived as more likely to be genuine. "Contextual embedding" means giving statements that place the event within the context of a specific time and place. Deceivers often leave out these connections to actual events. Therefore, contextual embedding increases the perception a review is likely genuine (Ansari & Gupta, 2021b). "Detailing" means providing sufficient details in a text or review so that only few open questions about the product remain. In many contexts, detailing is associated with genuineness, because it implies knowledge and experience with the product. However, in online reviews it is perceived as deceptive because consumers believe excessive details are suspicious (Ansari & Gupta, 2021b). "Flattering", on the other hand, refers to the existence of stylistic textual elements which review authors uses to appear pleasant and to connect with the audience. In an online review, flattering could be represented through the use of exclamation marks or emoticons. Flattering often leads to reviews being perceived as not genuine, because, for example the excessive use of exclamation marks, is assumed to reflect harmful intent in a reviewer (Ansari & Gupta, 2021b).

3.3.3. Reviewer characteristics

Another important aspect consumers look at when trying to identify the veracity of a review is the identity of the reviewer (Ansari et al., 2018; DeAndrea et al., 2018; Kusumasondjaja et al., 2012; Luo et al., 2013; Munzel, 2016; Peng et al., 2016; Román et al., 2019, pp. 141–166). That is, the more information there is about the reviewer, the more likely the review is perceived to be genuine. If the perceived similarity between the reader and the reviewer is high, the review is also more likely to be perceived as genuine (Racherla et al., 2012). Consumers judge the similarity between themselves and reviewers based on background information and details reviewers provide. For example, opinions or preferences expressed in a review that match those of the consumer, and sociodemographic background information that some platforms provide, such as age or nationality (Racherla et al., 2012; Román et al., 2019, pp. 141–166).

DeAndrea et al. (2018) found that the more independent from the product the reviewer is perceived to be then the more genuine is the review perceived to be, and the less influence the seller has on the displayed reviews the more genuine the reviews are perceived to be. Luo et al. (2013) found that a high source credibility can increase the recommendation credibility.

3.3.4. Seller characteristics

Peng et al. (2016) found that participants consider seller reputation and the trade record, the higher/better they are, the more trustworthy the reviews of the product are perceived to be. The more influence the seller has on the website the less trustworthy the reviews are perceived to be (Ansari et al., 2018; DeAndrea et al., 2018). This means that if it is perceived that the seller can delete or publish the displayed reviews (on their own business website for example) the reviews are perceived as less trustworthy.

3.3.5. Platform characteristics

Platform characteristics include not only which platform is used (company website or a website with different vendors) but also the layout, design and the information displayed on the platform (Ansari et al., 2018; Munzel, 2015). Different studies identified that it is important on what kind of website the review is displayed, for example on the company website, a platform for reviews or a neutral website (Ansari et al., 2018; DeAndrea et al., 2018; Filieri, 2016; Román et al., 2019, pp. 141–166). Furthermore, if the website displays fake reviews and marks them as such, the website and the reviews on that website are appraised as more trustworthy (Ananthakrishnan et al., 2020; Munzel, 2015, 2016). These studies show that participants place more trust in websites that display fraudulent reviews, and a summarized version of the fraud review information in the form of a trust score (Ananthakrishnan et al., 2020; Munzel, 2016). This suggests that consumers prefer to be aware of fraudulent reviews and that they likely prefer it when their cognitive load is reduced in the form of a summary score instead of having to go through all reviews themselves.

Munzel (2015) shows in an experiment that a third-party seal, indicating that the fake review detection algorithm used by a website is approved, and the display of information about the detection support mechanisms used by a website increase the perceived trustworthiness of reviews.

Three studies indicated that prior knowledge and training affects veracity decisions. DeAndrea et al., (2018), Munzel (2015), and Munzel (2016) show that knowledge about review spam has a positive effect on accurately detecting fake reviews. However, Kronrod et al. (2017) found that educating users about specific cues to deception to participants makes them overly suspicious. This means that participants tended to assume real reviews were fake. This could lead to the popularity and sales of platform sellers to decrease (Ansari et al., 2018). However, this negative effect of knowledge about cues to deception ceases to occur if a third-party seal suggests that the website uses a good detection support mechanism (Munzel, 2016).

3.4. Methods used

There are different approaches used across the reviewed studies. Four studies used a mixed-methods approach (Ansari et al., 2018; Ansari & Gupta, 2019, 2021b; Kronrod et al., 2017; Peng et al., 2016; Racherla et al., 2012). Two studies conducted interviews to gather more insight into how consumers deal with online consumer reviews (Filieri, 2016; Román et al., 2019, pp. 141–166) and seven studies conducted experiments (Ananthakrishnan et al., 2020; DeAndrea et al., 2018; Jensen et al., 2013, 2013, 2013; Kusumasondjaja et al., 2012; Munzel, 2015, 2016).

3.4.1. Mixed-methods

Four studies used a mixed-methods approach that combined qualitative and quantitative methods. Ansari et al. (Ansari et al., 2018; Ansari & Gupta, 2019, 2021b) used interviews as a preliminary study and followed these with an experimental survey and reported the combined findings across three different papers. Peng et al. (2016) conducted interviews first and then conducted an online survey without any experimental component. Kronrod et al. (2017) used a different approach. They asked participants to first generate fake and genuine reviews and then asked a second batch of participants to judge veracity of the generated reviews in an experimental survey. Racherla et al. (2012) first conducted focus groups with undergraduate students and then conducted an experimental survey with a second batch of participants.

Ansari et al. (Ansari et al., 2018; Ansari & Gupta, 2019, 2021b) conducted an exploratory study interviewing 20 Indian online shoppers to understand customers' information processing mechanisms and strategies adopted to deal with review manipulation in India. Peng et al. (2016) conducted 16 semi-structured interviews, asking participants about their knowledge and experience in fake review detection. Racherla et al. (2012) conducted focus groups with 20 undergraduate students while Kronrod et al. (2017) randomly assigned 1190 participants from the United States of America to one of six conditions to write review for a hotel or motel stay. In their authentic condition, participants wrote a review for a hotel they really experienced. In the fake condition they were assigned to one out of five groups. In the first group the participants were not given a clue on how to best write a fake review that seems genuine. In the other four conditions they were given the clues: to use more past tense, more unique words, more abstract language, and more personal pronouns.

In their main study, Ansari et al. (2018) randomly assigned 202 post-graduate students into one of four groups asking them to sort 30 reviews into fake and truthful reviews (Ansari et al., 2018; Ansari & Gupta, 2021b). Similar to this, Kronrod et al. (2017) asked 328 MTurk workers to each read and sort 60 reviews (30 real and 30 fake reviews from their study 1 control group where no directions were given with regard to how to successfully write a fake review) as either genuine or fake with five different conditions getting different (or no) clues beforehand. Peng et al. (2016), on the other hand, conducted a pen and paper survey with 199 Chinese students. Their survey aimed to assess how consumers perceive different manipulation tactics. Racherla et al. (2012) used a different approach conducting a 2 x 2 (within-subjects) x2 (between-subjects) experimental design to depict the factors that influence the consumers trust in a review (manipulated variables: valence and argument quality of the content and sociodemographic information). They asked 283 undergraduates to search the internet for a hotel to stay there with their friends. Then they were presented with four reviews and were asked to assign trust scores (Racherla et al., 2012).

3.4.2. Interviews

While in the mixed-method approaches interviews were conducted as exploratory pre-studies, two studies conducted interviews as the main study to gather more insight into how consumers deal with online consumer reviews (Filieri, 2016; Román et al., 2019, pp. 141–166). Filieri (2016) conducted 38 in-depth interviews with TripAdvisor users to ask about their experiences with online user reviews in the United Kingdom. Román et al. (2019, pp. 141–166) held 18 in-depth, semi--structured interviews where participants were asked to describe their most recent experiences with online consumer reviews before buying a product or service. After that, a few open-ended questions were asked, concerning how often they read reviews, on which platforms and their trust in reviews.

3.4.3. Experiments

While three studies conducted a single experimental survey (DeAndrea et al., 2018; Jensen et al., 2013; Kusumasondjaja et al., 2012), there were two studies that conducted multiple experiments in the form of surveys (Ananthakrishnan et al., 2020; Munzel, 2015). Munzel (2016) first conducted an analysis of real online reviews and then used the data to conduct an experimental survey with 390 participants. Lastly, Jensen et al. (2013) conducted a two-part experimental laboratory survey with 231 students.

DeAndrea et al. (2018) conducted a 2 (review spamming knowledge) × 2 (review platform) between-subjects experimental design where 123 participants were randomly assigned to one of four conditions in the United States of America. As the review spamming knowledge factor, the authors gave the participants one of two articles, one was an article about fake reviews on Yelp as an awareness stimulus and the second was an article about iPhones as a control stimulus. For the review platform factor, the authors created dummy Yelp webpages and dummy restaurant webpages for restaurants that do not exist (DeAndrea et al., 2018).

Jensen et al. (2013) conducted a 2 (lexical complexity: high versus low) x 2 (two-sidedness: high versus low) × 2 (affect intensity: high versus low) fully crossed experimental design in the United States of America. The 231 participants were randomly assigned into one of the

experimental conditions, which consisted of eight separate product webpages. There was one base product review which was edited for all combinations of high and low: (1) lexical complexity (2) two-sidedness (3) affect intensity (Jensen et al., 2013).

Kusumasondjaja et al. (2012) conducted a 2x2 between-subjects experiment with travellers from 31 different countries visiting Bali at the time of recruitment. The reviews that were displayed to the participants were manipulated on message valence (positive and negative) and the reviewers' identity (identified and anonymous) (Kusumasondjaja et al., 2012).

Ananthakrishnan et al. (2020) conducted three online experiments in the United States of America, where the participants selected a review from a fictive restaurant review portal. They had 820 participants in total (238 first; 293 s; 289 third). They measured different behaviors of the groups by different metrics such as clicks, time spent, number of page visits, and page activities in the first experiment. In the second experiment they measured the influence of: fraud information displayed, fraud information displayed together with a trust score,[7] only trust score displayed, and silent approach with a note that reviews were removed. They asked participants to make a bet on which review they trusted more; they were also able to divide their virtual chip. The third experiment tested how fake reviews influenced consumer behavior under different product quality settings (high, medium and low quality) by asking participants again to bet which restaurant would win a competition for being the best restaurant (all the needed information was available on the website); they could also divide their money (against and for the restaurant). After the bet was placed, participants were informed that the website contained fake reviews and were offered the opportunity to look at the fake reviews and change their bets (Ananthakrishnan et al., 2020).

Munzel (2015) conducted three online experiments in France with a total of 549 participants (first 197; second 211; third 141). In all experiments, the control group was given no additional information, while the other groups were primed by showing them a news article about the existence of fake reviews. In addition, three additional (dis-)trusting cues were tested: an independent Consumer Report Seal for detection capabilities of the website; the review site's own indication of the probability that a review is fake; or information about the overall rating of the product (Munzel, 2015).

In a follow up study, Munzel (2016) conducted a related 3 (identity disclosure) x2 (consensus) x2 (priming of fake reviews) experiment in France with 390 participants. They were confronted with newspaper articles that either activated fake review knowledge or not and were then exposed to a screenshot of a review for a restaurant. The review was manipulated to have either high, medium or low identity disclosure and were either in consensus with the overall rating or not. Afterwards, participants were asked about their perception of trustworthiness, purchase intention, and avoidance behavior (Munzel, 2016).

Luo et al. (2013) conducted an online questionnaire with 199 students in China to test whether recommendation persuasiveness affects the perceived credibility of a review. First participants were asked to read a new recommendation based on their current need or interest and copy the link into the questionnaire. The second part consisted of the questions for the constructs of their model and the third part were demographic questions (Luo et al., 2013).

3.5. Theories used in research on human fake review detection

In this section we introduce the theories used across the reviewed studies and describe how they were applied to consumers review veracity decisions. Warranting theory was used by three studies (Ansari et al., 2018; Ansari & Gupta, 2019; DeAndrea et al., 2018). The elaboration likelihood model (ELM) was used by three studies (Ansari & Gupta, 2019; Luo et al., 2013; Román et al., 2019, pp. 141–166), in all cases the ELM was used in combination with other theories. Two studies used uncertainty reduction theory (Kusumasondjaja et al., 2012; Racherla et al., 2012) and one study (Ansari & Gupta, 2021b) used speech act theory (SAT). Credibility theory was used by two studies (Filieri, 2016; Jensen et al., 2013). There were five studies that did not use existing theoretical frameworks (Ananthakrishnan et al., 2020; Kronrod et al., 2017; Munzel, 2015, 2016; Peng et al., 2016).Of these five studies, four based their research questions and hypotheses on earlier research without constructing a specific theoretical model (Ananthakrishnan et al., 2020; Kronrod et al., 2017; Munzel, 2015; Peng et al., 2016).

3.5.1. Warranting theory

Warranting theory was used by three studies (Ansari et al., 2018; Ansari & Gupta, 2019; DeAndrea et al., 2018), one of which combined it with the ELM (Ansari & Gupta, 2019). The key principle of the warranting theory is that the perception of information control reflects the impression of the information (Ansari et al., 2018; Ansari & Gupta, 2019; DeAndrea et al., 2018). This means that if the receiver of information has the impression that the sender has a strong influence on the information, the receiver will perceive the information as less valuable, meaning it has a low warranting value. In our context this means that the greater the manipulation of the content (e.g. the ability to edit or censor reviews) is perceived by the consumers, the lesser the authenticity of the review would be perceived to be (DeAndrea et al., 2018). The studies show that warranting theory can explain the relationship between the perception of a sender (e.g. a reviewer) of a message and the decision about how credible the message is perceived to be. Ansari et al. (2018) showed that review extremity, helpfulness, informativeness, subjectivity, and objectivity will be perceived as more (helpfulness voting, subjectivity, and moderate informativeness) or less (extreme star ratings, objectivity, and extreme or no informativeness) impressive and therefore affect the warranting value of the review. DeAndrea et al. (2018) showed that knowledge about review spamming, and review platform (if the reviews appear on a business review website or on the website of the business itself) influence the perceived veracity of a review by increasing or decreasing the warranting value. The more neutral and independent a review platform is perceived, the higher is the perceived veracity of the reviews (DeAndrea et al., 2018). Ansari et al. (2018) state in their article that the affect on the warranting value, as explained above, can also be explained by the ELM, and the ELM can additionally explain why some information is more persuasive than other information.

3.5.2. Elaboration Likelihood Model (ELM)

The ELM was used by three studies (Ansari & Gupta, 2019; Luo et al., 2013; Román et al., 2019, pp. 141–166). Of these studies, one combined it with the cognitive dissonance theory (Román et al., 2019, pp. 141–166), one combined it with warranting theory (Ansari & Gupta, 2019), and one combined it with theoretical considerations about information quality (Luo et al., 2013). The key principle of the ELM is that information is processed either via a central or a peripheral route (Petty & Brinol, 2011; Petty & Cacioppo, 1984). When the receiver of a message is motivated and able to actively process the information, it is processed via the central route and systematically analyzed in terms of the strength of arguments, credibility of included facts, and compatibility with prior knowledge to form a firm decision which is not easily changed afterwards. However, if the motivation or ability of the receiver is low, information is processed peripherally and heuristically analyzed. This then leads to a decision which can be more readily persuaded to change (Petty & Brinol, 2011; Petty & Cacioppo, 1984). Studies show that the use of the ELM can help identify factors that can influence the

[7] Since it needs a high cognitive engagement to process all the information, for each restaurant, a trust score was presented based on all reviews of the restaurant to indicate the proportion of fraudulent reviews. And it was displayed as a 'review quality score'.

identification of a review as either genuine or fake (Ansari et al., 2018; Román et al., 2019, pp. 141–166). Both Ansari et al. (2018) and Román et al. (2019, pp. 141–166) differentiate between central cues (Argument quality, source quality, and message length in Ansari et al. (2018); argument quality and review quantity in Román et al. (2019, pp. 141–166)) and peripheral cues (verified badges, most recent reviews, and 1 & 5 star ratings in Ansari et al. (2018); review quantity, review consistency and review homophily in Román et al. (2019, pp. 141–166)). Additionally, Román et al. (2019, pp. 141–166) identified moderator variables, such as purchase context, the review platform, hedonic versus utilitarian purchases and purchases involving negative needs versus positive wants. Other studies also show that the identity of a source (Peng et al., 2016) and prior knowledge about review spam can influence the perception of the veracity of a review (DeAndrea et al., 2018; Peng et al., 2016).

3.5.3. Cognitive dissonance theory

As mentioned above, one study combined Cognitive Dissonance Theory with the ELM (Román et al., 2019, pp. 141–166). A person has a cognitive state of attitudes, beliefs and behaviors and most of the time they will act accordingly. However, Festinger's Cognitive Dissonance Theory (Festinger, 1957) states that if one does something that contradicts this cognitive state, it will cause a dissonance, meaning a state of tension or arousal. This discomfort can lead to either a change in behavior or a change in attitude and therefore repair the dissonant cognitive state. Festinger argues that people are motivated to be consistent and avoid the state of dissonance even if that means changing their behavior or attitude (Festinger, 1957; Román et al., 2019, pp. 141–166). Román et al. (2019, pp. 141–166) used the ELM to explain how the consumer identifies the veracity of a review, while they used Cognitive Dissonance Theory to explain what happens when the review is identified as a fake positive review, with the level of cognitive dissonance provoked by the discovery a review was false motivating increasingly severe responses. If the customer thinks the deception is minor, they will simply not purchase the product. When the deception is perceived to be stronger they will report the review, and if the deception is perceived to be especially strong they may even stop buying anything from the company altogether (Román et al., 2019, pp. 141–166).

3.5.4. Uncertainty reduction theory

Two studies used Uncertainty Reduction Theory (Kusumasondjaja et al., 2012; Racherla et al., 2012) which states that the moment two strangers meet, the primary concern of both is to reduce the uncertainty or increase the predictability of the behavior of all parties involved in the interaction (Racherla et al., 2012). It further states that uncertainty has an influence on the level of intimacy, liking and trust between the parties involved and has three phases (1) entry phase, (2) personal phase and (3) attitude assessment phase. In the context of reviews this means that since there is no history and no future expectations of interaction between the sender and receiver of an online consumer review, active and passive strategies are used to evaluate the information (Racherla et al., 2012). Active strategies could be scanning other reviews of the same product or the review itself for arguments, and passive strategies could be comparing socio-demographic characteristics of reviewers. Racherla et at. (2012) found that high similarity and high argument quality both increase the trust in a review, similar to Ansari et al. (2018) and Román et al. (Román et al., 2019, pp. 141–166) who found argument quality to be a central cue in fake review detection using the Elaboration Likelihood Model. Kusumasondjaja et al. (2012) on the other hand found that the identity of the reviewer is of high importance. When the reviewer is unknown uncertainty is high and so the review is perceived as less trustworthy, while trustworthiness increases when there is more information about the reviewer.

3.5.5. Speech act theory

Speech Act Theory was used by one study (Ansari & Gupta, 2021b) and proposes that next to the actual statements that are said or written and understood by a sender and receiver, respectively, there is a second message in spoken and written language, called an utterance (Ansari & Gupta, 2021b). These utterances describe an indicative action or request a form of action. Speech Act Theory further states that there are three different messages in a statement: the statement itself, what the sender intends to convey with the message, and the effect of the statement on the receiver (Ansari & Gupta, 2021b). An example of this would be: the traffic light is green. At the first level it describes the fact that the traffic light is green. At the second level, the statement may be intended as a request to drive on. At the third level, the receiver could interpret the statement as a sign that the sender believes that the receiver is inattentive and so needs a signal to indicate that they should proceed. Ansari and Gupta (Ansari & Gupta, 2021b) provide insight that the perceived veracity of a review can be influenced by referencing, flattering, contextual embedding, detailing and argument structure. Referencing, for example using pronouns, can be used to connect with a reader which influences the receivers' comprehension, in this study, they found that the use of more pronouns is perceived as more deceptive (Ansari & Gupta, 2021b). Detailing can help to support decisions or recommendations, however, as mentioned earlier, in reviews it is perceived as suspicious. Flattering can be understood as emojis, quotation and exclamation marks. In online reviews they can build rapport and help convey the emotions of the reviewer to the receiver. However, in reviews they are perceived to be used as manipulation, using flattery to influence the receiver to feel a certain way about the product being reviewed (Ansari & Gupta, 2021b). Contextual embedding indicates a personal experience, and helps the sender to connect to the receiver, showing them the whole picture. Too little contextual embedding can therefore create distance and indicate deception (Ansari & Gupta, 2021b). Argument structure, measured here with cognitive process words, can indicate a thoughtfulness and therefore, show that the sender wants to make sure something is understood correctly, which in online reviews is perceived as less likely to be deceptive.

3.5.6. Credibility theory

Credibility Theory was used by two studies (Filieri, 2016; Jensen et al., 2013). Credibility Theory states that credibility results from the interaction of source characteristics (such as expertise, reputation and labels), characteristics of the message (argument quality and plausibility), characteristics of the receiver (own cultural background, involvement, beliefs and motivation) and the media via which the message is conveyed (design features such as usability and ease of navigation) (Filieri, 2016; Wathen & Burkell, 2002). However, the process of how these factors are processed is not clearly described by the theory. Similar to Elaboration Likelihood Model and Speech Act theory, Credibility Theory takes into account argument quality as factor. The theory assumes that a higher argument quality has a positive effect on the credibility of a review. While Jensen et al. (2013) combined Credibility with Language Expectancy Theory, Filieri (2016) mainly used Credibility Theory to develop their interview guide.

3.5.7. Language Expectancy Theory

Jensen et al. (2013) used Credibility Theory in combination with Language Expectancy Theory. Language Expectancy Theory describes the effect of language characteristics from different groups of individuals on the change of attitudes (Burgoon & Miller, 2018; Jensen et al., 2013). It includes that when we know a person has a certain characteristic or profession, we expect them to communicate in a certain way. For example, we expect more formal communication from a professor than from a student. It also includes how persuasive the communication is expected to be (Burgoon & Miller, 2018; Jensen et al., 2013). Following Jensen et al. (2013), this means one expects to read within reviews a certain level of language complexity; one-sidedness of arguments with extreme ratings, and two-sidedness with moderate ratings. If these expectations are violated, the perceived credibility of the

reviewer and the perception of the product are affected negatively (Jensen et al., 2013).

3.5.8. Other/ no theoretical background

Munzel (2016) used earlier research from different authors to build a new theoretical framework. He stated that identity disclosure, consensus information and persuasion knowledge activation influence the trustworthiness of a review, which in turn influences the purchase intention and avoidance intention. Munzel (2016) furthermore states that consensus information has a moderating effect on the identity disclosure of a reviewer, which means that if there is little information on the identity of the reviewer but high consensus between reviewers is high, then the trust is still high. The theory argues that the more is known about the sender/reviewer the more trustworthy the message/review is perceived to be.

Román et al. (2019) used the findings of their interviews to develop a new theoretical framework. The interviews were developed by considering a combination of Cognitive Dissonance Theory and the Elaboration Likelihood Model. The new theoretical framework described by Román et al. (2019, pp. 141–166) proposes that elaboration motivation and elaboration ability influence the central (argument quality and review quantity) and peripheral cues (review quantity, review consistency and review homophily) to veracity, while product knowledge acts as a moderator on the strength of these relationships. The central and peripheral cues and initial expectations then influence the **p**erceived **d**eception in **o**nline **c**onsumer **r**eviews (PDOCR). Perceived deception in turn influences self-protective and public and revenge behaviors. Roman et a. (2019) also find that there are moderating variables that have an influence on the perceived deception and the resulting behaviour behavior. There were purchase context (products vs. services), review platform (retailer's website/other platforms), hedonic vs. utilitarian purchases, and purchases involving negative needs vs. positive wants).

Next to using an existing theoretical framework, Credibility Theory, Filieri (2016) used his findings to develop a new theoretical framework. Filieri argued that the inductively generated cues he identified (source trustworthiness, message trustworthiness, review valence and pattern in reviews) determined consumers' perception of trustworthiness of a review. This trustworthiness in turn determined how persuasive a review was. Filieri (2016) also found that consumer involvement, consumer experience with consumer reviews and the medium type are moderating factors in determining how cues impacted on trustworthiness, and via trustworthiness, persuasion.

There were four studies that based their research questions and hypotheses on earlier research without constructing a specific theoretical model (Ananthakrishnan et al., 2020; Kronrod et al., 2017; Munzel, 2015; Peng et al., 2016).

4. Discussion: A research agenda for consumers' fake review detection

This review gives an overview of the cues and features used by consumers to detect fake reviews in online shopping systems. We found 15 papers that identified or tested different cues used by consumers to detect the veracity of a review. These cues have been identified within a diversity of theoretical and methodological approaches. Overall, we show that in contrast to algorithmic fake detection, human fake review detection is a multi-faceted phenomenon in which situational as well as cognitive elements play a role. The most common elements identified in literature are that argument quality, information about reviewers' identity, and the characteristics of the reviewer (what the reviewer is like) have a positive influence on how genuine a review is perceived to be. Furthermore, many theories indicate that how the (fake) reviews are displayed, including on which website, has an influence on how genuine reviews are perceived to be. However, few theories are able to both describe the cues and factors that are used by consumers to identify the veracity of an online review, and also describe the process by which the cues and factors identified lead to decisions by consumers.

We will first discuss the strengths and limitations of the theories and methods captured within our review. Afterwards, we compare human detection and automated detection, discuss the limitations of our literature review and then discuss future work.

4.1. Strengths and limitations of the theories applied to fake review detection

We found very little consistency across the literature with regard to theoretical approach. Consequently, we review the strengths and limitation of each theory. The theoretical frameworks used were Warranting theory, Elaboration Likelihood Model, Cognitive Dissonance Theory, Uncertainty Reduction Theory, Speech Act Theory, Credibility Theory and Language Expectancy Theory.

When comparing these theories and how they are used in human fake review detection online, we found that Language Expectancy Theory, Uncertainty Reduction Theory and Credibility Theory propose a mental model of what consumers believe an online consumer review ought to be like. In simplified terms, these theories suggest that people perceive online reviews as suspicious if the reviews do not match these mental models. Similarly, Warranting Theory proposes that people are more likely to believe a message when they trust the sender. Warranting Theory has thereby been used to explain how the consumers' perception of the author of a review influences how they perceive the review. The Elaboration Likelihood Model (ELM) allows us to understand how consumer motivation and ability leads to (sub)optimal decision making depending on whether consumers are willing or able to process online reviews via the central or peripheral route. Whereas the previous theories explain how people try to detect fake reviews, Cognitive Dissonance Theory and Speech Act Theory explain how the identification of fake online reviews influences the buying behavior of consumers.

As a consequence of these strengths and weaknesses, we argue that a combination of the Language Expectancy Theory, Uncertainty Reduction Theory, Credibility Theory, Elaboration Likelihood Model, Cognitive Dissonance Theory and Speech Act Theory would be most valuable in explaining what review cues are used by humans and how they are considered during the act of purchasing.

There are also various limitations of the application of theory in our identified literature to be mentioned. Firstly, on a theoretical level, none of the theories explain sufficiently how the different cues/factors are considered during the purchase and how much they weigh into the decision-making process. Furthermore, it is also noteworthy that most studies employed a deductive approach using a priori (categories of) cues presumed to be relevant from prior research, rather than using an inductive approach. Since most of the theories applied deductively are taken from adjacent research areas, a lack of inductive research comes at the risk of overlooking important cues that are specific to fake review detection. This concern is compounded by the scarcity of research on this topic overall. There is a real need to galvanize research on this topic to ensure relevant cues are identified. Additionally, even where studies were consistent in using the same or similar theoretical approaches, they often used different cues and therefore it is hard to determine how to combine the theories and cues most effectively. We hope that our identification of specific clusters of cues will help to make it easier for future work in this area to be both more comprehensive in their consideration of cues, but also more specific with regard to how these cues lead to behaviors.

On an empirical level, we find that the research on how consumers identify the veracity of online consumer reviews is very scarce. This leads us to consider that there might be more cues that have not been included in the research and theories yet. Coupled with the uncertainty we identify regarding how cues are processed, we conclude that it is not yet possible based on the extant literature to develop a full-fledged theoretical framework explaining how consumers detect fake online consumer reviews. Developing such a theoretical framework therefore is

a priority for future work examining fake review detection. Based on the findings of our study the foundations of such a theoretical framework should integrate the best qualities of the theories that identify the relevant the cues, and those of the theories that focus on identifying how consumers process information. We advocate for more inductive work to developing theory in this area to complement the more extensive deductive work that has already been done. This would allow any new theory to be grounded in observations of actual consumer behavior, complementing the existing theories we review here.

4.2. Strengths and limitations of the methods applied in fake review detection

From the 13 studies in our literature review (three papers from Ansari describe one mixed-methods design) there were four mixed-methods approaches (Ansari et al., 2018; Kronrod et al., 2017; Peng et al., 2016; Racherla et al., 2012) and two studies that conducted qualitative interviews (Filieri, 2016; Román et al., 2019, pp. 141–166). Though mixed-methods approaches offer the possibility to first develop a theoretical framework and then test it, none of the above-mentioned studies did so. The mixed-method approaches all used preexisting theoretical frameworks and tested the influence of specific aspects on the perceived veracity of a review. The qualitative parts of the mixed-method designs were primarily descriptive rather than theoretical. That is, they gathered insight into the awareness of fake reviews of consumers and asked which cues consumers used (Ansari et al., 2018; Ansari & Gupta, 2019, 2021b; Peng et al., 2016) or they asked participants to produce online consumer reviews which would then be used in follow up studies (Kronrod et al., 2017). However, there was no attempt to generate and test a new theory. As a result, these studies are less capable of depicting the whole picture.

Explorative qualitative interviews are a good method to explore how a decision about the veracity of a review is made, and so allow an inductive analysis of processes or mechanisms that underpin decision making. However, Filieri (2016) and Román et al. (2019, pp. 141–166) both constructed their interview guide based on existing theoretical frameworks. Due to the use of existing and limited theoretical frameworks they may have impeded participants from describing novel processes or cues that have not previously been considered in the fake review detection literature. Furthermore, both Román et al. (2019, pp. 141–166) and Filieri (2016) embed the cues they found almost completely within the Elaboration Likelihood Model by disambiguating the found cues into central versus peripheral cues. In later articles, Filieri et al. (Filieri, Hofacker, & Alguezaui, 2018; Filieri, Raguseo, & Vitari, 2018) highlight that the primary benefit of their approach was the inductive identification of relevant cues, rather than developing new theory to explain how those cues are processed.

The consequence of the at least partly deductive approach to the extant qualitative literature we identify, is that there is still a need for exploratory qualitative research to uncover cues and relations not covered by existing theories. There remains a need to use methods that will help us to develop theory for understanding fake review detection.

The non-qualitative studies used an experimental design to test specific factors. The strength of these studies is that they can empirically test some of the theoretical assumptions outlined above, such as argument quality and consensus. However, in such an underdeveloped field of research as human fake review detection there remains significant risk that many key variables are unknown, and where they are known we cannot yet be confident in how they affect judgments. Again, we argue that research that builds theory based on observation of behavior and through exploring the thought processes of actual consumers would complement the existing evidence base and allow greater confidence when describing relevant cues and how they are processed. In this way we can better determine which theory or theories best describe how consumers make veracity judgements about online reviews, and how these are then translated into concrete actions.

4.3. Comparing human and automated fake review detection

When comparing human versus automated fake review detection, there are cues that are used by both, such as review characteristics (Ansari & Gupta, 2021b; Heydari et al., 2015) or reviewer characteristics (DeAndrea et al., 2018). Examples of review characteristics are usefulness rating counts (Khurshid et al., 2019), linguistic features (Ott et al., 2011; Reddy et al., 2020) and metrics such as word counts (Anderson & Simester, 2014; Dewang, Singh, & Singh, 2016). Reviewer characteristic are features that can identify the reviewer such as a name (Barbado et al., 2019), but also the number of friends or followers (Z. Wu et al., 2017). Furthermore, both humans and machines use textual characteristics, however, while with humans textual characteristic are content based such as the amount of detail (Ansari et al., 2018; Ansari & Gupta, 2021b; Jensen et al., 2013; Kronrod et al., 2017; Román et al., 2019, pp. 141–166), machine learning focuses on the amount of adverbs, verbs or nouns used, the percentage of capitals used or on other text statistics (Fontanarava et al., 2017; Jindal & Liu, 2008).

However, there are also differences between the human and machine approach. Unlike machine approaches, humans also consider seller characteristics, such as trade records (DeAndrea et al., 2018; Peng et al., 2016; Racherla et al., 2012) and characteristics of the platform where the review is displayed (Ananthakrishnan et al., 2020; Munzel, 2015, 2016). On the other hand, machine fake review detection incorporates product centric cues such as price (Jindal & Liu, 2008), and meta-feature centric features such as date variance (Fontanarava et al., 2017). The reason for these discrepancies between users and automated fake review detection could be that product centric features such as the price of the item are perceived by consumers as less affected by reviews. However, the reasons underpinning why different cues are used by consumers has not yet been fully explored which is an additional motivator for exploratory research that addresses such questions.

In addition to differences in the cues used, humans and machines process the information differently. For instance, theories such as the ELM argue that humans process information by two different routes (the central and the peripheral one) (Ansari et al., 2018), but we do not know of any algorithmic approach that processes the information in the same way.

4.4. Key lessons and suggestions for future research based on the current findings

Our first key lesson is that because of the diversity of theories and cues employed, it remains unclear how well the cues and cognitive processes involved in fake review detection have been mapped by the extant literature. For example, the theoretical framework incorporating the Elaboration Likelihood Model and Cognitive Dissonance theory by Román (Román et al., 2019, pp. 141–166) includes only some of the factors identified as relevant by the studies we have analyzed in this systematic review; and does not include factors that influence the trustworthiness and credibility of reviews. We argue that more freely developed interview guidelines could help develop a new theoretical framework that does cover all factors, cues, prioritizing and thought processes in identifying the veracity of a review. In other words, there is a need for more inductive bottom-up research based on the lived experiences of consumers to supplement the predominantly deductive top-down approach that has been taken so far.

Second, during our research we identified two approaches to human detection of fake reviews which complemented automated detection: (1) cues to deception and (2) research on trustworthiness and credibility of reviews. This is a meaningful discovery because it shows that human and automated fake review detection are not interchangeable. This is further supported by research on automated review detection which showed that automated review detection outperforms humans (Ott et al., 2011) and research which shows that if human detection is supported by automated detection, the detection rate increases (Kim et al., 2021;

Yuan et al., 2016). Furthermore, research on automated fake review detection shows that there is no tool that reaches absolute accuracy (Ahmed et al., 2018; Jindal & Liu, 2008; Kennedy et al., 2019; Ott et al., 2011). Therefore, we propose that research should focus on researching both areas thoroughly and combining them for higher accuracy in fake review detection as 'human-in-the-loop'. The new generation of fake reviews are a serious problem for detection if one concentrates solely on textual features, because they are often written by humans who were instructed on how to write reviews (Bode, 2022; Stiftung Warentest, 2020), therefore a different approach has to be developed (Hovy, 2016).

To summarize, further research should address the limitations of the current theoretical approaches by combining their stronger elements with a new inductively generated theory which would be developed into interventions that are focused on increasing the salience of cues intuitively attended to by humans as well as those that can be trained. Likely these would include cues that are already successfully used by machine learning approaches. Any such theory needs to be more than just a collection of cues and must also explain how and why people attend to relevant cues, and how we can transfer the perception of cues to improved behavior.

Additionally, we do not differentiate between different industries, such as hotels, products or services, within the review. We expect it to be likely that the nature (and cost) of the purchased item or experience may influence consumer behavior and judgements. Unfortunately, we do not believe it is currently possible to make reliable claims about how reviews have different effects depending on industry or item due to the scarcity of the research. We do, however, recommend using different industries as a predictor variable in future research.

Concretely, a research agenda to address the research gaps we identify might begin with qualitative work to understand how consumers themselves believe they think about and determine the veracity of a review, including what cues they attend to and how these are combined with additional considerations. For example, it seems likely that consumers prior beliefs about a product, brand or platform are likely to influence purchases. As are decision heuristics such as confirmation bias. These findings can then be integrated with the theoretical frameworks. This combined inductive and deductive approach would provide a solid foundation for a comprehensive theoretical framework. This new framework should then be tested and further developed both in the laboratory and on actual consumer behavior.

4.5. Strengths and limitations of this review

Identifying that there is a limited evidence base is one contribution of this review, and we hope that this review will provide a springboard for future research. Nonetheless, the limitations of this study are partly determined by the scarcity of and inconsistency in the extant research. This is exemplified by the use of multiple different words to describe fake reviews, including but not limited to 'opinion spam', 'opinion scam', 'opinion fraud', 'review spam' and 'online review'. Furthermore, there are different words used for cues including but not limited to 'features', 'factors', 'indicators' and 'detection strategies'. We addressed this variety of terms by extending our initial backward and forward search for potential keywords, and also included many terms in our search string. It might however be the case that certain key words were omitted from our search meaning that we cannot guarantee that we included all research there is on consumer fake review detection.

Similarly, as for all systematic reviews, new research is continually published after the search is complete. In an effort to ensure our article is as up to date as possible we have also conducted a post-review search. This search identified an additional six articles on fake review detection which have been omitted from the initial search, however none of them fit our inclusion criteria. Five of them discuss algorithmic ways to identify fake reviews online (Birim et al., 2022; Lee, Song, Li, Lee, & Yang, 2022; Salminen, Kandpal, Kamel, Jung, & Jansen, 2022; Shi et al., 2022; Tufail et al., 2022). The sixth recent study investigated the consequences of perceived credibility of exaggerated positive online consumer reviews (Román et al., 2023). The authors found that perceived review credibility affected the consumers thoughts about the brand's reputation, purchase intention, and perceptions about trustworthiness of the review site. In turn, brand identification enhanced review credibility (Román et al., 2023). However, since Román et al. (2023) do not discuss any detection cues, this article also does not meet our inclusion criteria.

A related limitation is that we considered only peer reviewed literature and did not seek grey literature on this topic. Therefore, we excluded student projects and one PhD project (Roland, 2019) that would otherwise have fit the other criteria of our review study. Nonetheless, we argue that we found a variety of different research that has to be combined in future research. That is, we believe we have captured the breadth of the literature even if it is never possible to have identified the full depth of extant research. Capturing the breadth rather than depth of the research is more fundamental for our research goals however, since we aimed to consider the diversity of cues, theories and methods used in the published literature. Furthermore, we argue that the holistic way we have evaluated the existing literature by considering the strength of the theory and methods and not only looking at outcomes, is another strength of this literature review.

5. Conclusion

Online fake reviews are a challenge for online markets and platforms, since they undermine user ratings as a tool for consumers to make an informed choice. Given the shortcomings of machine learning approaches to identify fake reviews, it stands to reason that there is benefit in supporting consumers in identifying fake reviews themselves. However, consumer detection cues for online fake reviews are not yet well understood. Our systematic review has consolidated the scattered research landscape by analyzing existing cues used by consumers as well as research on factors that influence the trustworthiness and credibility of online reviews. Our survey shows that these are a broad variety of cues and factors which can be categorized into five categories:

(1) review characteristics
(2) textual characteristics
(3) reviewer characteristics
(4) seller characteristics
(5) characteristics of the platform where the review is displayed.

However, since research on this topic is scarce it is possible that there are yet undiscovered factors and cues. Hence, it should be researched which cues consumers believe are most relevant and how they influence consumer behavior. This requires the development of a consolidated theory that exploits the best aspects of extant theories, but which is also supplemented by theory grounded in the experiences of consumers.

Consent for publication

All authors and the University of Applied Sciences Bonn-Rhein-Sieg consent in the publication of this article. This manuscript has not been published in any other journal and is not under review by any other journal.

Author contributions

The authors contributions are as follows, the first author wrote most of the article as part of her PhD program at the University of Applied Sciences Bonn-Rhein-Sieg and the University of Twente. The supervisors, Prof. Dr. Gunnar Stevens and Dr. Timo Jakobi, were involved in the design idea and throughout the whole project, the editing process and approval decision to publish the article. Dr. Steven Watson was involved in co-drafting the article, the editing process and approval decision to

publish the article.

Funding

None.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Appendix A

Key words: opinion fraud, fake reviews, opinion spam, detection cues, recommender systems, user-perspective, detection strategies, persuasive recommendations, consumer, Opinion fraud detection, online deception, online review, opinion spam detection, opinion spam features, review spam, deception detection, spam indicators, trust, lie detection.

Appendix B

Research direction	Study	Theory	Method	Cues
Factors that influence perceived veracity inductive and deductive	Ansari S, Gupta S, Dewangan J.Do Customers Perceive Reviews as Manipulated? A Warranting Theory Perspective. 2018; 16. (Ansari et al., 2018)	Warranting Theory + review characteristics & textual characteristics	Mixed-method approach, first exploratory interviews 20 online shoppers Main study: survey to measure impression of an online review (veracity), we randomly selected 120 reviews from the review corpus of 34,389 reviews. We then divided these 120 reviews into 4 blocks containing 30 reviews each. Each block was evenly randomly presented to the survey participants such that each participant receives a block i.e. 30 reviews around 50 respondents per group (202 useable in total)	Dependent: impression of a review explanatory: review extremity, helpfulness, informativeness, subjectivity, and objectivity Writing style: Pronouns and adverbs Amount of detail, spelling and grammar
	Kronrod A, Lee JK, Gordeliy I.Detecting Fictitious Consumer Reviews: A Theory-Driven Approach Combining Automated Text Analysis and Experimental Design. Mark Sci Inst Work Pap Ser. 2017; 17–124. (Kronrod et al., 2017)	Linguistic Characteristics of Lies Characteristics of authentic vs fictitious descriptions of experiences	Mixed-method approach Study 1: Generating reviews, either real or fake, In fake they were given clues: no clue (control), past tense, unique words, abstract language, personal pronouns Study 2: Asking participants to determine for 60 reviews the veracity, using the same clues as in study 1 328 useable MTurk workers 60 reviews were 30 real and 30 fake reviews from study 1 (control group) 5 conditions: no clue, past tense, unique words, abstract language, personal pronouns	Study 2: Most commonly mentioned: Avoidance of detail and specific descriptions or names, Vagueness and omission of information, Overly short or long reviews, Extreme valence (overly, negative/ positive/good/bad), Lack of descriptions of personal experience, The presence of grammar/spelling errors
	Ansari, S., & Gupta, S. (2021b). Customer perception of the deceptiveness of online product reviews: A speech act theory perspective. *International Journal of Information Management, 57*, 102286. https://doi.org/10.1016/j.ijinfomgt.2020.102286 (Ansari & Gupta, 2021b)	Speech act theory	Survey 120 reviews randomly picked from flipkart about phones, 4 blocks of each 30 reviews, each participant only 1 block, 202 responses in total Text mining the reviews with LIWC for independent variables (referencing, contextual embedding, detailing, flattering and argument structure) dependent variable: fake or genuine Fractional logit model	Referencing Flattering Contextual embedding Detailing Argument structure
	Ansari, S., & Gupta, S. (2019). FAKE REVIEWS AND MANIPULATION: DO CUSTOMER REVIEWS MATTER? Research Papers. https://aisel.aisnet.org/ecis2019_rp/143 (Ansari & Gupta, 2019)	Elaboration Likelihood Model & warranting theory	Qualitative interviews to understand customers' information processing mechanism and strategies adopted to deal with review manipulation	Do not mention the cues because it was not their goal = review assessment Central cues: Argument quality, source quality, message length Peripheral cues: verified badge, most recent reviews, 1& 5 star ratings
	DeAndrea DC, Van Der Heide B, Vendemia MA, Vang MH.How People Evaluate Online reviews.	Warranting theory	Evaluation of information online? 2 × 2 between- subjects experimental design, deductive	review spamming knowledge (article about spam vs control) review platform (the reviews appeared on a business review website (i.e., Yelp.

(*continued on next page*)

(*continued*)

Research direction	Study	Theory	Method	Cues
	Commun Res. 2018 Jul 1; 45(5):719–36. DeAndrea et al. (2018)			com). In the other condition, the same reviews, originating from Yelp.com, appeared on the restaurant's own website.)
	Peng L, Cui G, Zhuang M, Li C.Consumer perceptions of online review deceptions: an empirical study in China. J Consum Mark. 2016 Jan 1; 33(4):269–80. (Peng et al., 2016)	Earlier research without specific theoretical framework Deceptive practices in marketing Study 2: Information manipulation theory (IMT) research finds	Mixed-methods Study 1: 16 Interviews Study 2: Survey Objective: how do participants perceive different manipulation tactics? To ensure consistent understanding of the tactics, participants were presented the description of each of the three manipulation tactics, at random order to avoid bias. Participants were then asked how they perceived the tactics in terms of deceptiveness, ease of detection and ethicality and indicate their purchase intention and perceived usefulness of online product reviews.	Study 1: the quality of content; the quantity of the reviews; the mismatch between reviews, seller reputation and the trade record; and the identities of the reviewers
	Román S, P. Riquelme I, Iacobucci D. Perceived Deception in Online Consumer Reviews: Antecedents, Consequences, and Moderators. In 2019. p. 141–66. (Román et al., 2019, pp. 141–166)	Elaboration Likelihood Model and Cognitive Dissonance Theory	In-depth interviews, inductive	New theoretical framework: central cues (argument quality and review quantity), peripheral cues (review quantity, review consistency and review homophily), Moderatoor variables: purchase context, the review platform, hedonic versus utilitarian purchase and purchase involving negative needs versus positive want
Factors that influence perceived trustworthiness	Ananthakrishnan UM, Li B, Smith MD.A Tangled Web: Should Online Review Portals Display Fraudulent Reviews? Inf Syst Res. 2020 Sep; 31(3):950–71. (Ananthakrishnan et al., 2020)	Earlier research without specific theoretical framework	Three randomized user experiments, deductive	Displaying fake reviews
	Filieri R.What makes an online consumer review trustworthy? Ann Tour Res. 2016 May 1; 58:46–64. (Filieri, 2016)	Credibility Theory	Interviews & grounded theory approach, inductive	New theoretical framework: the content and writing style of a review message, the valence and review extremity, the source of communication, the pattern in reviews, receivers' involvement, Consumer experience, Medium, Consumer decision making influence perceived trustworthiness which influences persuasion
	Munzel A. Malicious practice of fake reviews: Experimental insight into the potential of contextual indicators in assisting consumers to detect deceptive opinion spam. Rech Appl En Mark Engl Ed. 2015 Dec 1; 30(4):24–50. (Munzel, 2015)	Earlier research without specific theoretical framework	Three experimental studies, deductive	Priming, consensus, displaying of fake reviews, explanation about detection support mechanism
	Munzel A.Assisting consumers in detecting fake reviews: The role of identity information disclosure and consensus. J Retail Consum Serv. 2016 Sep 1; 32:96–108.(Munzel, 2016)	Earlier research with own theoretical framework	3 × 2 × 2 between-subjects factorial design, deductive	identity-descriptive information, consensus, persuasion knowledge
	Racherla P, Mandviwalla M, Connolly DJ.Factors affecting consumers' trust in online product reviews. J Consum Behav. 2012; 11(2):94–104. (Racherla et al., 2012)	Uncertainty reduction theory	Mixed-methods, focus grous and 2 × 2 × 2 full factorial experimental design, deductive	Argument quality, perceived similarity between subject and reviewer
Factors that influence perceived credibility	Jensen M, Averbeck J, Zhang Z, Wright K.Credibility of Anonymous Online Product Reviews: A Language Expectancy Perspective. J Manag Inf Syst. 2013 Jul 1; 30:293–324. (Jensen et al., 2013)	Language Expectancy Theory and Credibility theory	2 × 2 × 2 fully crossed experimental design, deductive	lexical complexity, two-sidedness and affect intensity
	Kusumasondjaja S, Shanka T, Marchegiani C.Credibility of online reviews and initial trust: The roles of reviewer's identity and review valence. J Vacat Mark. 2012 Jul 1; 18(3):185–95. (Kusumasondjaja et al., 2012)	Uncertainty reduction Theory	2x2 between subject experimental, deductive	message valence and reviewer's identity

(*continued on next page*)

(*continued*)

Research direction	Study	Theory	Method	Cues
	Luo C, Luo XR, Schatzberg L, Sia CL. Impact of informational factors on online recommendation credibility: The moderating role of source credibility. Decis Support Syst. 2013; 56:92–102. (Luo et al., 2013)	Elaboration Likelihood Model and information quality	Questionnaire, deductive	recommendation persuasiveness, recommendation completeness' and recommendation credibility

Appendix C

Review centric features (Ahmed et al., 2018; Algur, Patil, Hiremath, & Shivashankar, 2010; Alsubari et al., 2020, pp. 3846–3856; Anderson & Simester, 2014; Banerjee & Chua, 2014; Barbado et al., 2019; Budhi et al., 2021; Dewang et al., 2016; Fontanarava et al., 2017; Harris, 2018; Jindal & Liu, 2008; Khurshid et al., 2019; Kotian & Meshram, 2017; Y. Li, Feng, & Zhang, 2016; Ong, 2013; Ott et al., 2011; Pillala et al., 2020; Pokharkar et al., 2017, p. 7; Reddy et al., 2020; Runa, Zhang, & Zhai, 2017; Saini, Verma, & Sharan, 2019; Shah & Ahmed, 2019; Shehnepoor et al., 2017; Shukla et al., 2019; Z. Wu et al., 2017; Yao et al., 2017).	Rating features	Useful rating count given by other consumers	Khurshid et al., 2019 (Khurshid et al., 2019)
	Linguistic features Fontanarava et al., 2017; Ott et al., 2011; Pokharkar et al., 2017, p. 7; Reddy et al., 2020; Shehnepoor et al., 2017; Shukla et al., 2019; Yao et al., 2017 (Fontanarava et al., 2017; Ott et al., 2011; Pokharkar et al., 2017, p. 7; Reddy et al., 2020; Shehnepoor et al., 2017; Shukla et al., 2019; Yao et al., 2017)	use of adverbs	Ahmed et al., 2018; Alsubari et al., 2020, pp. 3846–3856; Banerjee & Chua, 2014; Budhi et al., 2021; Dewang et al., 2016; Ott et al., 2011 (Ahmed et al., 2018; Alsubari et al., 2020, pp. 3846–3856; Banerjee & Chua, 2014; Budhi et al., 2021; Dewang et al., 2016; Ott et al., 2011)
		adjectives and verbs	Ahmed et al., 2018; Alsubari et al., 2020, pp. 3846–3856; Banerjee & Chua, 2014; Budhi et al., 2021; Dewang et al., 2016; Ott et al., 2011 (Ahmed et al., 2018; Alsubari et al., 2020, pp. 3846–3856; Banerjee & Chua, 2014; Budhi et al., 2021; Dewang et al., 2016; Ott et al., 2011)
		word count	Alsubari et al., 2020; Anderson & Simester, 2014; Dewang et al., 2016; Fontanarava et al., 2017; Khurshid et al., 2019; Reddy et al., 2020; Saini et al., 2019; Wu et al., 2017 (Alsubari et al., 2020, pp. 3846–3856; Anderson & Simester, 2014; Dewang et al., 2016; Fontanarava et al., 2017; Khurshid et al., 2019; Reddy et al., 2020; Saini et al., 2019; Z. Wu et al., 2017)
		number of characters	Dewang et al., 2016; Khurshid et al., 2019 (Dewang et al., 2016; Khurshid et al., 2019)
		number of sentences	Dewang et al., 2016; Saini et al., 2019 (Dewang et al., 2016; Saini et al., 2019)
		duplicates	Algur et al., 2010; Jindal & Liu, 2008; Pillala et al., 2020 (Algur et al., 2010; Jindal & Liu, 2008; Pillala et al., 2020)
Reviewer centric	Trusting features Barbado et al., 2019; Fazzolari et al., 2020; Goswami et al., 2017; Jindal & Liu, 2008 (Barbado et al., 2019; Fazzolari et al., 2020; Goswami et al., 2017; Jindal & Liu, 2008)	number of photos	Barbado et al., 2019; Fazzolari et al., 2020; Goswami et al., 2017 (Barbado et al., 2019; Fazzolari et al., 2020; Goswami et al., 2017)
	Social features Barbado et al., 2019; Goswami et al., 2017; Wang et al., 2016; Wu et al., 2017 (Barbado et al., 2019; Goswami et al., 2017; Wang et al., 2016; Z. Wu et al., 2017)	number of friends	Barbado et al., 2019; Goswami et al., 2017; Wang et al., 2016; Wu et al., 2017 (Barbado et al., 2019; Goswami et al., 2017; Wang et al., 2016; Z. Wu et al., 2017)
	Personal features Barbado et al., 2019; Goswami et al., 2017; Wang et al., 2016; Wu et al., 2017 (Barbado et al., 2019; Goswami et al., 2017; Wang et al., 2016; Z. Wu et al., 2017)	real name	Barbado et al., 2019; Pillala et al., 2020 (Barbado et al., 2019; Pillala et al., 2020)
	Activity features (Barbado et al., 2019; Fazzolari et al., 2020; Goswami et al., 2017; Harris, 2018) Barbado et al., 2019; Fazzolari et al., 2020; Goswami et al., 2017; Harris, 2018	number of reviews wrote	Barbado et al., 2019; Fazzolari et al., 2020; Harris, 2018 (Barbado et al., 2019; Fazzolari et al., 2020; Harris, 2018)
		Inter-reviews similarity	Gera et al., 2015; Jindal & Liu, 2008 (Gera et al., 2015; Jindal & Liu, 2008),
	Reviewer centric linguistic and behavioral pattern Pokharkar et al., 2017; Shehnepoor et al., 2017 (Pokharkar et al., 2017, p. 7; Shehnepoor et al., 2017)	Same reviewer	Le et al., 2022 (Le et al., 2022)
Meta-data features	Temporal features Fontanarava et al., 2017; Khurshid et al., 2019 (Fontanarava et al., 2017; Khurshid et al., 2019)	date variance	Fontanarava et al., 2017 (Fontanarava et al., 2017)
	Ratings Fontanarava et al., 2017; Jindal & Liu, 2008; Khurshid et al., 2019; Pillala et al., 2020; Runa et al., 2017 (Fontanarava et al., 2017; Jindal & Liu, 2008; Khurshid et al., 2019; Pillala et al., 2020; Runa et al., 2017)	Rating deviation from average rating score of the product	Fontanarava et al., 2017; Jindal & Liu, 2008; Runa et al., 2017 (Fontanarava et al., 2017; Jindal & Liu, 2008; Runa et al., 2017)
		Rating score given by the review	Jindal & Liu, 2008; Wang et al., 2016 (Jindal & Liu, 2008; Wang et al., 2016)

(*continued on next page*)

(*continued*)

Product centric features	Jindal & Liu, 2008; Wang et al., 2016 (Jindal & Liu, 2008; Wang et al., 2016)	Price	Jindal & Liu, 2008 (Jindal & Liu, 2008)
		Average rating score of the product	Jindal & Liu, 2008 (Jindal & Liu, 2008)

References

Ahmed, H., Traore, I., & Saad, S. (2018). Detecting opinion spams and fake news using text classification. *Security and Privacy, 1*(1), e9. https://doi.org/10.1002/spy2.9

Algur, S. P., Patil, A. P., Hiremath, P. S., & Shivashankar, S. (2010). Conceptual level similarity measure based review spam detection. In *2010 international conference on signal and image processing* (pp. 416–423).

Alsubari, S., Shelke, M., & Deshmukh, S. (2020). *Fake reviews identification based on deep computational linguistic features*.

Amblee, N., & Bui, T. (2011). Harnessing the influence of social proof in online shopping: The effect of electronic word of mouth on sales of digital microproducts. *International Journal of Electronic Commerce, 16*(2), 91–114. https://doi.org/10.2753/JEC1086-4415160205

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. https://doi.org/10.1287/isre.2020.0925

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. https://doi.org/10.1509/jmr.13.0209

Ansari, S., & Gupta, S. (2019). *Fake reviews and manipulation: DO customer reviews matter? Research papers*. https://aisel.aisnet.org/ecis2019_rp/143.

Ansari, S., & Gupta, S. (2021a). Review manipulation: Literature review, and future research agenda. *Pacific Asia Journal of the Association for Information Systems, 13*(1). https://doi.org/10.17705/1pais.13104

Ansari, S., & Gupta, S. (2021b). Customer perception of the deceptiveness of online product reviews: A speech act theory perspective. *International Journal of Information Management, 57*, Article 102286. https://doi.org/10.1016/j.ijinfomgt.2020.102286

Ansari, S., Gupta, S., & Dewangan, J. (2018). *Do Customers Perceive Reviews as Manipulated? A Warranting Theory Perspective, 16*.

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. https://doi.org/10.1016/j.jretconser.2013.03.004

Banerjee, S., & Chua, A. Y. K. (2014). Dissecting genuine and deceptive kudos: The case of online hotel reviews. *International Journal of Advanced Computer Science and Applications*, 28–35.

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. https://doi.org/10.1016/j.ipm.2019.03.002

Biolchini, J., Mian, P. G., Natali, A. C. C., & Travassos, G. H. (2005). *Systematic review in software engineering* (Vol. 679, p. 45). System Engineering and Computer Science Department COPPE/UFRJ, Technical Report ES, 05.

Birim, Ş.Ö., Kazancoglu, I., Kumar Mangla, S., Kahraman, A., Kumar, S., & Kazancoglu, Y. (2022). Detecting fake reviews through topic modelling. *Journal of Business Research, 149*, 884–900. https://doi.org/10.1016/j.jbusres.2022.05.081

Bode, B. (2022). *Fake Reviews auf Amazon: Nun klagt der Online-Konzern. BASIC thinking*. https://www.basicthinking.de/blog/2022/02/22/fake-reviews-auf-amazon-nun-klagt-der-online-konzern/.

BSI. (2022). *Digitalbarometer—Bürgerbefragung zur Cyber-Sicherheit 2022 [Kurzbericht zur Studie der Polizeilichen Kriminalprävention der Länder und des Bundes (ProPK) und des Bundesamtes für Sicherheit in der Informationstechnik (BSI)]. Polizeilichen Kriminalprävention der Länder und des Bundes (ProPK) undBundesamt für Sicherheit in der Informationstechnik*. https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Digitalbarometer/Digitalbarometer-ProPK-BSI_2022.pdf?_blob=publicationFile&v=3.

Budhi, G. S., Chiong, R., & Wang, Z. (2021). Resampling imbalanced data to detect fake reviews using machine learning classifiers and textual-based features. *Multimedia Tools and Applications, 80*(9), 13079–13097. https://doi.org/10.1007/s11042-020-10299-5

Burgoon, M., & Miller, G. (2018). An expectancy interpretation of language and persuasion. In H. Giles, & R. N. S. Clair (Eds.), *Recent advances in language, communication, and social Psychology* (pp. 199–225). Routledge.

Button, M., & Cross, C. (2017). *Cyber frauds, scams and their victims*. Routledge. https://doi.org/10.4324/9781315679877

Button, M., Nicholls, C. M., Kerr, J., & Owen, R. (2014). Online frauds: Learning from victims why they fall for these scams. *Australian and New Zealand Journal of Criminology, 47*(3), 391–408.

Castelli, N., Stevens, G., & Jakobi, T. (2019). *Information visualization at home: A literature survey of consumption feedback design*.

Chevalier, S. (2021). *Online shopping frauds UK 2020*. Statista. https://www.statista.com/statistics/1182221/online-shopping-fraud-incidents-uk/.

Cohen, J. (2020). *How to spot a fake review on Amazon*. PCMag UK. https://uk.pcmag.com/old-shopping/121470/how-to-spot-a-fake-review-on-amazon.

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. https://doi.org/10.1177/0093650215573862

Dewang, R. K., Singh, P., & Singh, A. K. (2016). Finding of review spam through "corleone, review genre, writing style and review text detail features. In *Proceedings of the second international conference on information and communication technology for competitive strategies* (pp. 1–6). https://doi.org/10.1145/2905055.2905081

Ezhilarasan, M., Govindasamy, V., Akila, V., & Vadivelan, K. (2019). Sentiment analysis on product review: A survey. In *2019 international Conference on Computation of power, energy, Information and communication (ICCPEIC)* (pp. 180–192). https://doi.org/10.1109/ICCPEIC45300.2019.9082346

Fazzolari, M., Buccafurri, F., Lax, G., & Petrocchi, M. (2020). *Improving opinion spam detection by cumulative relative frequency distribution. ArXiv:2012.13905 [Cs]* http://arxiv.org/abs/2012.13905.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.

Filieri, R. (2016). What makes an online consumer review trustworthy? *Annals of Tourism Research, 58*, 46–64. https://doi.org/10.1016/j.annals.2015.12.019

Filieri, R., Hofacker, C. F., & Alguezaui, S. (2018). What makes information in online consumer reviews diagnostic over time? The role of review relevancy, factuality, currency, source credibility and ranking score. *Computers in Human Behavior, 80*, 122–131. https://doi.org/10.1016/j.chb.2017.10.039

Filieri, R., Raguseo, E., & Vitari, C. (2018). When are extreme ratings more helpful? Empirical evidence on the moderating effects of review characteristics and product type. *Computers in Human Behavior, 88*, 134–142. https://doi.org/10.1016/j.chb.2018.05.042

Fontanarava, J., Pasi, G., & Viviani, M. (2017). Feature analysis for fake review detection through supervised classification. In *2017 IEEE international conference on data science and advanced Analytics (DSAA)* (pp. 658–666). https://doi.org/10.1109/DSAA.2017.51

Gera, T., Thakur, D., & Singh, J. (2015). BILD testing for spotting out suspicious reviews, suspicious reviewers and group spammers. In *2015 fifth international Conference on communication Systems and network technologies* (pp. 976–981). https://doi.org/10.1109/CSNT.2015.138

Goswami, K., Park, Y., & Song, C. (2017). Impact of reviewer social interaction on online consumer review fraud detection. *Journal of Big Data, 4*(1), 1–19.

Gupta, M., & George, J. (2013). The Effects of Fictitious Customer Testimonials across Different Cultures. In *AMCIS 2013 Proceedings*. https://aisel.aisnet.org/amcis2013/HumanComputerInteraction/RoundTablePresentations/12.

Harris, C. G. (2012). *Detecting deceptive opinion spam using human computation* (Vol. 7).

Harris, C. G. (2018). Decomposing TripAdvisor: Detecting potentially fraudulent hotel reviews in the era of big data. In *2018 IEEE international Conference on big knowledge (ICBK)* (pp. 243–251). https://doi.org/10.1109/ICBK.2018.00040

Harzing, A. W.. Publish or Perish. available from https://harzing.com/resources/publish-or-perish.

Hernandez, N., Rahman, M., Recabarren, R., & Carbunar, B. (2018). Fraud De-anonymization for fun and profit. In *Proceedings of the 2018 ACM SIGSAC conference on computer and communications security* (pp. 115–130). https://doi.org/10.1145/3243734.3243770

Heydari, A., Tavakoli, M., & Salim, N. (2015). *A framework for review spam detection research*.

Heydari, A., Tavakoli, M., & Salim, N. (2016). Detection of fake opinions using time series. *Expert Systems with Applications, 58*, 83–92. https://doi.org/10.1016/j.eswa.2016.03.020

Higgins, J., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M., et al. (2021). *Cochrane Handbook for systematic Reviews of interventions. Cochrane training*. https://training.cochrane.org/handbook/current.

Hossain, M. A., Noor, R. M., Yau, K.-L. A., Azzuhri, S. R., Z'aba, M. R., & Ahmedy, I. (2020). Comprehensive survey of machine learning approaches in cognitive radio-based vehicular ad hoc networks. *IEEE Access, 8*, 78054–78108. https://doi.org/10.1109/ACCESS.2020.2989870

Hovy, D. (2016). 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* (Vol. 2, pp. 351–356). https://doi.org/10.18653/v1/P16-2057. *Short Papers)*.

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.

Hussain, N., Turab Mirza, H., Rasool, G., Hussain, I., & Kaleem, M. (2019). Spam review detection techniques: A systematic literature review. *Applied Sciences, 9*(5). https://doi.org/10.3390/app9050987. Article 5.

Jansen, J., & Leukfeldt, R. (2018). Coping with cybercrime victimization: An exploratory study into impact and change. *Journal of Qualitative Criminal Justice and Criminology, 6*(2), 205–228.

Jensen, M., Averbeck, J., Zhang, Z., & Wright, K. (2013). Credibility of anonymous online product reviews: A Language Expectancy perspective. *Journal of Management Information Systems, 30*, 293–324. https://doi.org/10.2753/MIS0742-1222300109

Jindal, N., & Liu, B. (2007). Review spam detection. In *Proceedings of the 16th international conference on world wide Web* (pp. 1189–1190). https://doi.org/10.1145/1242572.1242759

Jindal, N., & Liu, B. (2008). Opinion spam and analysis. In *Proceedings of the 2008 international conference on Web search and data mining* (pp. 219–230). https://doi.org/10.1145/1341531.1341560

Kennedy, S., Walsh, N., Sloka, K., Foster, J., & McCarren, A. (2019). Fact or factitious? Contextualized opinion spam detection. In *Proceedings of the 57th annual meeting of the association for computational linguistics: Student research workshop* (pp. 344–350). https://doi.org/10.18653/v1/P19-2048

Khurshid, F., Zhu, Y., Xu, Z., Ahmad, M., & Ahmad, M. (2019). Enactment of ensemble learning for review spam detection on selected features. *International Journal of Computational Intelligence Systems, 12*(1), 387–394. https://doi.org/10.2991/ijcis.2019.125905655

Kim, J., Kang, J., Shin, S., & Myaeng, S.-H. (2021). Can you distinguish truthful from fake reviews? User analysis and assistance tool for fake review detection. In *Proceedings of the first workshop on bridging human–computer interaction and natural language processing* (pp. 53–59). https://www.aclweb.org/anthology/2021.hcinlp-1.9.

Kitchenham, B. (2004). *Procedures for performing systematic reviews* (Vol. 33, pp. 1–26). Keele, UK: Keele University, 2004.

Korteling, J. E., Hans,)., van de Boer-Visschedijk, G. C., Blankendaal, R. A. M., Boonekamp, R. C., & Eikelboom, A. R. (2021). Human- versus artificial intelligence. In *Frontiers in artificial intelligence* (Vol. 4). https://www.frontiersin.org/articles/10.3389/frai.2021.622364.

Kotian, H., & Meshram, B. B. (2017). Detection of spam reviews and spammers in E-commerce sites. In *2017 international Conference on current Trends in computer, electrical, Electronics and communication (CTCEEC)* (pp. 299–304).

Kronrod, A., Lee, J. K., & Gordeliy, I. (2017). Detecting fictitious consumer reviews: A theory-driven approach combining automated text analysis and experimental design. *Marketing Science Institute Working Papers Series*, 17–124.

Kusumasondjaja, S., Shanka, T., & Marchegiani, C. (2012). Credibility of online reviews and initial trust: The roles of reviewer's identity and review valence. *Journal of Vacation Marketing, 18*(3), 185–195. https://doi.org/10.1177/1356766712449365

Lee, M., Song, Y. H., Li, L., Lee, K. Y., & Yang, S.-B. (2022). Detecting fake reviews with supervised machine learning algorithms. *Service Industries Journal, 42*(13–14), 1101–1121. https://doi.org/10.1080/02642069.2022.2054996

Le, T.-K.-H., Li, Y.-Z., & Li, S.-T. (2022). Do reviewers' words and behaviors help detect fake online reviews and spammers? Evidence from a hierarchical model. *IEEE Access, 10*, 42181–42197. https://doi.org/10.1109/ACCESS.2022.3167511

Li, H., Fei, G., Wang, S., Liu, B., Shao, W., Mukherjee, A., et al. (2016a). *Modeling review spam using temporal patterns and Co-bursting behaviors. ArXiv:1611.06625 [Cs]* http://arxiv.org/abs/1611.06625.

Li, Y., Feng, X., & Zhang, S. (2016b). Detecting fake reviews utilizing semantic and emotion model. In *2016 3rd international conference on information science and control engineering (ICISCE)* (pp. 317–320). https://doi.org/10.1109/ICISCE.2016.77

Luo, C., Luo, X. R., Schatzberg, L., & Sia, C. L. (2013). Impact of informational factors on online recommendation credibility: The moderating role of source credibility. *Decision Support Systems, 56*, 92–102.

Malbon, J. (2013). Taking fake online consumer reviews seriously. *Journal of Consumer Policy, 36*(2), 139–157. https://doi.org/10.1007/s10603-012-9216-7

Masip, J., Bethencourt, M., Lucas, G., Segundo, M. S.-S., & Herrero, C. (2012). Deception detection from written accounts. *Scandinavian Journal of Psychology, 53*(2), 103–111. https://doi.org/10.1111/j.1467-9450.2011.00931.x

Mayzlin, D., Dover, Y., & Chevalier, J. (2014). Promotional reviews: An empirical investigation of online review manipulation. *The American Economic Review, 104*(8), 2421–2455. https://doi.org/10.1257/aer.104.8.2421

Mohawesh, R., Xu, S., Tran, S. N., Ollington, R., Springer, M., Jararweh, Y., et al. (2021). Fake reviews detection: A survey. *IEEE Access, 9*, 65771–65802. https://doi.org/10.1109/ACCESS.2021.3075573

Munzel, A. (2015). Malicious practice of fake reviews: Experimental insight into the potential of contextual indicators in assisting consumers to detect deceptive opinion spam. *Recherche et Applications en Marketing, 30*(4), 24–50. https://doi.org/10.1177/2051570715604155

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. https://doi.org/10.1016/j.jretconser.2016.06.002

Ong, T. C. (2013). Product reputation manipulation: The characteristics and impact of shill reviews [ph.D., university of Colorado at denver]. In ProQuest dissertations and theses. https://www.proquest.com/docview/1399161414/abstract/8FA9E4D6B14949B6PQ/1.

Ott, M., Choi, Y., Cardie, C., & Hancock, J. T. (2011). *Finding deceptive opinion spam by any stretch of the imagination. ArXiv:1107.4557 [Cs]* http://arxiv.org/abs/1107.4557.

Patel, N. A., & Patel, R. (2018). A survey on fake review detection using machine learning techniques. In *2018 4th international Conference on computing Communication and automation (ICCCA)* (pp. 1–6). https://doi.org/10.1109/CCAA.2018.8777594

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. https://doi.org/10.1108/JCM-01-2015-1281

Petty, R. E., & Brinol, P. (2011). The elaboration likelihood model. In P. A. M. V. Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social Psychology: Collection: Volumes 1 & 2* (pp. 224–245). SAGE.

Petty, R. E., & Cacioppo, J. T. (1984). *Source factors and the elaboration likelihood model of persuasion. ACR North American Advances, NA-11* https://www.acrwebsite.org/volumes/6328/volumes/v11/NA-11/full.

Pillala, P., Bhansali, P. K., Reddy, M. A. E., & Rojamani, G. (2020). Random forest model for intrusion detection in crowd-sourced reviews. In *2020 international Conference on smart Electronics and communication (ICOSEC)* (pp. 325–329). https://doi.org/10.1109/ICOSEC49089.2020.9215423

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. https://doi.org/10.1016/j.jbusres.2018.12.009

Pokharkar, S. T., Shete, A. J., & Ghogare, V. D. (2017). *Survey in Online social media Skelton by network based spam*, 04(11.

Racherla, P., Mandviwalla, M., & Connolly, D. J. (2012). Factors affecting consumers' trust in online product reviews. *Journal of Consumer Behaviour, 11*(2), 94–104. https://doi.org/10.1002/cb.385

Rastogi. (2020). *Effective opinion spam detection: A study on review metadata versus content.* http://manu47.magtech.com.cn/Jwk3_jdis/EN/10.2478/jdis-2020-0013.

Rastogi, A., & Mehrotra, M. (2017). Opinion spam detection in online reviews. *Journal of Information and Knowledge Management*. https://doi.org/10.1142/S0219649217500368

Reddy, O. V. C., Babu, S. R., Scholar, P. G., & Tech, M. (2020). *A novel supervised and semi-supervised learning based fake online reviews detection system* (Vol. 5), 0886.

Ren, Y., & Ji, D. (2019). Learning to detect deceptive opinion spam: A survey. *IEEE Access, 7*, 42934–42945. https://doi.org/10.1109/ACCESS.2019.2908495

Roland, C. R. (2019). *Online deceit:the Use of idiosyncratic Cues in identifying duplicitous user-generated content [Thesis*. Purdue University Graduate School. https://doi.org/10.25394/PGS.8968673.v1

Román, S., Riquelme, I. P., & Iacobucci, D. (2019). *Perceived deception in online consumer reviews: Antecedents, consequences, and moderators*. https://doi.org/10.1108/S1548-643520190000016010

Román, S., Riquelme, I. P., & Iacobucci, D. (2023). Fake or credible? Antecedents and consequences of perceived credibility in exaggerated online reviews. *Journal of Business Research, 156*, Article 113466. https://doi.org/10.1016/j.jbusres.2022.113466

Runa, D., Zhang, X., & Zhai, Y. (2017). Try to find fake reviews with semantic and relational discovery. In *2017 13th international Conference on semantics, Knowledge and grids (SKG)* (pp. 234–239). https://doi.org/10.1109/SKG.2017.00048

Saini, M., & Sharan, A. (2017). Identifying deceptive opinion spam using aspect-based emotions and human behavior modeling. *International Journal of Hospitality Information Technology, 10*(1), 447–456. https://doi.org/10.14257/ijhit.2017.10.1.36

Saini, M., Verma, S., & Sharan, A. (2019). Multi-view ensemble learning using rough set based feature ranking for opinion spam detection. In S. K. Bhatia, S. Tiwari, K. K. Mishra, & M. C. Trivedi (Eds.), *Advances in computer communication and computational Sciences* (pp. 3–12). Springer. https://doi.org/10.1007/978-981-13-0341-8_1.

Salminen, J., Kandpal, C., Kamel, A. M., Jung, S., & Jansen, B. J. (2022). Creating and detecting fake reviews of online products. *Journal of Retailing and Consumer Services, 64*, Article 102771. https://doi.org/10.1016/j.jretconser.2021.102771

Santos, A. S. dos, Camargo, L. F. R., & Lacerda, D. P. (2020). Evaluation of classification techniques for identifying fake reviews about products and services on the internet. *Gestão & Produção, 27*. https://doi.org/10.1590/0104-530X4672-20

Shah, F., & Ahmed, S. (2019). Fake review detection using principal component analysis and active learning. *International Journal of Computer Applications, 178*, 42–48. https://doi.org/10.5120/ijca2019919418

Shehnepoor, S., Salehi, M., Farahbakhsh, R., & Crespi, N. (2017). NetSpam: A network-based spam detection framework for reviews in online social media. *IEEE Transactions on Information Forensics and Security, 12*(7), 1585–1595. https://doi.org/10.1109/TIFS.2017.2675361

Shi, L., Xie, S., Wei, L., Tao, Y., Junaid, A. W., & Gao, Y. (2022). *Joint sentiment topic model with word embeddings for fake review detection*. https://doi.org/10.2139/ssrn.4096565. SSRN Scholarly Paper No. 4096565).

Shukla, A., Wang, W., Gao, G., Gordon), & Agarwal, R. (2019). *Catch me if you can—detecting fraudulent online Reviews of doctors using deep learning (SSRN scholarly paper ID 3320258). Social science research network*. https://doi.org/10.2139/ssrn.3320258

Spalek, B. (2016). *Crime victims: Theory, policy and practice*. Macmillan International Higher Education.

Spithoven, R. (2020). *Verbonden risico's—maatschappelijke veiligheid in de black box society. Boomcriminologie*.

Statista Research Department. (2021). E-Commerce. In *Deutschland: Daten und Fakten zum boomenden Onlinegeschäft*. Statista Research Department. https://de.statista.com/themen/247/e-commerce/.

Stevens, G., Boden, A., Alizadeh, F., Jakobi, T., Walther, M., & Krüger, J. (2023). Wie gehen verbraucher:innen mit onlinebetrug um? – Eine literaturübersicht. In T.-G. Rüdiger, & P. S. Bayerl (Eds.), *Handbuch cyberkriminologie* (Vol. 2, pp. 1–22). Springer Fachmedien. https://doi.org/10.1007/978-3-658-35450-3_42-1.

Trivedi, V. (2021). E-commerce fraud to surpass $20B in 2021, an 18% jump over last year, report finds. Payments Dive. https://www.paymentsdive.com/news/e-commerce-fraud-to-hit-20-billion-2021-an-18-jump-from-prior-year/599312/.

Tufail, H., Ashraf, M. U., Alsubhi, K., & Aljahdali, H. M. (2022). The effect of fake reviews on e-commerce during and after covid-19 pandemic: SKL-based fake reviews detection. *IEEE Access, 10*, 25555–25564. https://doi.org/10.1109/ACCESS.2022.3152806

Van Nek, L., & Bolz, C. (2021). *Digitalbarometer 2021: Bürgerbefragung zur Cyber-Sicherheit. Polizeilichen Kriminalprävention der Länder und des Bundes und Bundesamt für Sicherheit in der Informationstechnik*. https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Digitalbarometer/Digitalbarometer-ProPK-BSI_2021.html?nn=520690.

Vidanagama, D. U., Silva, T. P., & Karunananda, A. S. (2020). Deceptive consumer review detection: A survey. *Artificial Intelligence Review, 53*(2), 1323–1352. https://doi.org/10.1007/s10462-019-09697-5

Wang, X., Liu, K., He, S., & Zhao, J. (2016). Learning to represent review with tensor decomposition for spam detection. In *Proceedings of the 2016 conference on empirical methods in natural language processing* (pp. 866–875). https://doi.org/10.18653/v1/D16-1083

Warentest, S. (2020). *Fake-Bewertungen: Wie Verkäufer mit gekauftem Lob Kunden manipulieren.* Stiftung Warentest; Stiftung Warentest. https://www.test.de/Fake-Bewertungen-Wie-Verkaeufer-mit-gekauftem-Lob-Kunden-manipulieren-5401497-0/.

Wathen, C. N., & Burkell, J. (2002). *Believe it or not: Factors influencing credibility on the Web*. Journal of the American Society for Information Science and Technology.

Wu, Z., Cao, J., Wang, Y., Wang, Y., Zhang, L., & Wu, J. (2020a). hPSD: A hybrid PU-Learning-Based spammer detection model for product reviews. *IEEE Transactions on Cybernetics, 50*(4), 1595–1606. https://doi.org/10.1109/TCYB.2018.2877161

Wu, Y., Ngai, E. W. T., Wu, P., & Wu, C. (2020b). Fake online reviews: Literature review, synthesis, and directions for future research. *Decision Support Systems, 132*, Article 113280. https://doi.org/10.1016/j.dss.2020.113280

Wu, Z., Zhang, L., Wang, Y., & Cao, J. (2017). *Identifying spam in reviews*.

Yao, Y., Viswanath, B., Cryan, J., Zheng, H., & Zhao, B. Y. (2017). Automated crowdturfing attacks and defenses in online review systems. In *Proceedings of the 2017 ACM SIGSAC conference on computer and communications security* (pp. 1143–1158). https://doi.org/10.1145/3133956.3133990

Yuan, Y., Xie, S., Lu, C.-T., Tang, J., & Yu, P. S. (2016). Interpretable and effective opinion spam detection via temporal patterns mining across websites. In *2016 IEEE international Conference on big data (big data)* (pp. 96–105). https://doi.org/10.1109/BigData.2016.7840593

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. https://doi.org/10.1080/07421222.2016.1205907

Michelle Walther is a researcher at the University of Applied Sciences Bonn-Rhein-Sieg. She got her Bachelors and Masters Degree in Psychology of Conflict, Risk and Safety at the University of Twente (The Netherlands). During her Master Thesis she specialized in Conflicts, Risks and Safety and specifically on the detection of lies in the digital world. With this knowledge it is her goal to investigate how consumers recognize fake online reviews and how the resilience of consumers can be increased.

Timo Jakobi is a Post-Doc and managing director of the institute for digital consumption at the University of Applied Sciences Bonn-Rhein-Sieg. His main research direction lies in the investigation and development of useable data security and data safety solutions which can be used by companies within customer experience design. Within this Timo operates at a crossing of user research and jurisprudence, to facilitate a robust, empirical perspective of consumers on the definition and interpretation of (data security)laws. Topics include but are not restricted to the implementation of complex demands, such as transparency, self-explanatory intelligent systems, rights of data subjects and design of consent, taking into account consumer and company interests.

Steven Watson is an Assistant Professor within the Psychology of Conflict, Risk and Safety section at the University of Twente. Steven researches decision making and communication in applied contexts, including in the areas of deception detection and cybersecurity.

Gunnar Stevens is professor for system informatics, especially IT-Security at the University of Siegen (Germany) and co-director of the institute for digital consumption at the University of Applied Sciences Bonn-Rhein-Sieg. His research field includes but is not restricted to User-Centered Security, Useable Privacy & Data Literacy, as well as Consumer Data Analytics & Visualization. His expertise lies in informatics and Human-Computer-Interaction as well as the ergonomic design of data protection and information system solutions.