Erasmus University Rotterdam

Erasmus School of Economics

MSc. Economics & Business - Marketing

Beyond the Mask: Investigating the Role of Reviewer Anonymity in Shaping Online Consumer Content and Purchase Intention

Student Name: Carolina Rueda

Student ID: 687807

Supervisor: Dr. AC Bayerl

Date: 30/06/24

Abstract

This study investigates the effect of reviewer anonymity on content creation and content consumption within the framework of online reviews, addressing a significant gap in understanding how anonymity influences these processes. The first part covers the anonymity of the review writer and the effect when producing online reviews. The second part studies the anonymity of the review writer and its impact on perceived reliability and purchase intention. Employing a dual approach, this research combines big data analysis and experimental design. The content creation analysis utilizes secondary data from Amazon's online marketplace, encompassing 13.3 million reviews across 29 product categories from 2008 to 2018. To assess the impact on content consumption, a 2x2 within-subject design experiment was conducted, with 135 qualified responses included in the final analysis. The findings reveal that anonymous reviewers tend to give higher rating scores but shorter and more positive reviews than authentic reviewers. In terms of content consumption, the anonymity effect on purchase intention was found to be mediated by perceived reliability. Moreover, positive sentiment enhances perceived reliability, yet this effect is less pronounced for authentic reviews. While both authenticity and positive sentiment individually boost purchase intention, their combination produces a less harmonious impact than anticipated.

Keywords: online reviews, anonymity, review lenght, sentiment analysis, purchase intention

Contents

Table of Figures

Table of Tables

Table of Equations

1. Introduction

In the modern era, the internet has become a fundamental player in shaping the information landscape available to consumers. The internet serves as a repository of knowledge, offering wide access to information about products and facilitating consumers to make informed decisions. Among the variety of information sources, online reviews emerge as a key component, providing firsthand insights and experiences shared by fellow consumers. These reviews act as electronic word-of-mouth (eWOM), influencing purchasing decisions by offering authentic perspectives on products and services. As consumers navigate the digital marketplace, online reviews stand out as crucial guides, helping individuals filter through the excess of options and make choices aligned with their preferences and expectations.

Online reviews hold a significant influence over consumer decision-making processes. As a fundamental component of the online marketplace, reviews serve as a crucial source of information for potential customers, shaping perceptions of products and services (Zhang et al., 2019). They act as a virtual word-of-mouth, allowing consumers to glean insights from the experiences of others, ultimately steering their purchasing decisions and contributing to the dynamic landscape of e-commerce.

Overall, online reviews could affect companies on two main fronts: reviews as a cue for potential customers (Zhang et al., 2023) and a measure of customer satisfaction. From one point of view, as consumers increasingly rely on the experiences of others shared through these reviews, a positive or negative representation can significantly influence perceptions of a brand and influence consumers' purchase decisions. From another point of view, reviewing customer satisfaction metrics could allow companies to identify emerging trends and adapt to meet evolving customer expectations. Overall, understanding the dynamics of online reviews is key for businesses seeking to establish and maintain a healthy online reputation while increasing customer engagement.

The exploration of reviewer anonymity within the context of online reviews is a key point when understanding purchasing decisions. Anonymity increases the level of complexity of the evaluation of online feedback, as it raises questions about the credibility and authenticity of the reviewer's insights. Anonymity (or lack of identity disclosure) refers to the practice of users providing feedback without revealing their true identities. This anonymity can influence the nature and authenticity of the information disclosed in reviews and the trust levels accredited by potential customers. On one hand, it empowers users to express honest opinions without fear of personal repercussions. On the other hand, the lack of accountability may also lead to compromising the reliability of the overall feedback for potential buyers (Pu et al., 2020).

Thus, this paper aims to answer the following research questions: (1) *Does the anonymity of reviewers result in variations in the rating score, sentiment, and length of the reviews?* (2) *To what extent does the anonymity of a reviewer impact the perceived reliability of the review and affect purchase intention?*

Extent literature has been done on analyzing the impact of anonymity on online reviews from different perspectives. However, almost all the previous studies have focused on experimental designs to set the anonymous set up for participants and analyze their reactions. This paper is intended to provide an evaluation of the effect of anonymity using online reviews, instead of experimental designs, from Amazon and classifying usernames on anonymous vs. non-anonymous reviewers' names. Additionally, this study presents an analysis from two points of view: how anonymity could change user-generated content in online reviews and, secondly, how reviewers' anonymity name shapes the perceived trust and purchase intention of potential customers. This double-sided approach, from the researcher's knowledge, has not yet been done, and the results of this paper will enrich the current academic literature.

Furthermore, for managerial contributions, the target managers of this study are mainly customer insights managers who want to identify factors that could influence the ultimate purchase intention process of customers. Understanding to what extent the anonymity of reviewers might harm the perceived reliability by potential customers allows managers to work on initiatives to incentivize the right user-generated content. Increasing sales is the goal for companies; therefore, determining actions to increase conversion rates is vital, and external validity, given by online reviews, has been proven to play a key role (Zhang et al., 2023). Moreover, this study aims to provide insights into possible features such as anonymity on online marketplaces and how to target review content in favor of the brand image.

Therefore, this paper evaluates the impact of anonymity from two perspectives: content creation evaluated on review length, sentiment, and rating score, and content consumption through purchase intentions while emphasizing the roles of perceived reliability as a mediator and review sentiment as a moderator. By delving into the complex interplay between these factors, this research aims to provide a more comprehensive understanding of the key drivers behind the effect of anonymity in the context of electronic word of mouth (eWOM).

2. Theoretical Background and Research Hypotheses

With the ongoing increase in internet usage, online content has become one of the most direct and impactful ways to engage with consumers. As Rainie (2005) mentioned, "*The internet is constantly reshaping people's informational and social universes, but people are constantly reshaping the internet as well*" (p. 69), highlighting the importance of understanding this dynamic interaction, businesses need to account for new trends and how they influence consumer behavior, shaping purchase intention and determining its success.

Online content comes in various forms, each playing a crucial role in different marketing strategies. From blogs and articles to social media, online content is essential for effectively reaching targeted audiences and driving customer engagement and loyalty. Social media content, for example, helps increase engagement levels and build a sense of community among users. Blogs and articles provide detailed information, assisting brands to connect with audiences seeking expertise. Infographics, visuals, and videos simplify complex content for audiences with a preference for visual media. Lastly, user-generated content, such as reviews, are perceived as more credible since the company or seller is not directly involved in the content creation process (Filieri et al., 2021).

This research paper involves a comprehensive examination of anonymity in online reviews, addressing both the standpoint of content creators and potential buyers. This dual perspective enables a thorough comparison of findings, ensuring consistency in our analysis.

2.1 Online Reviews

Online reviews, as voluntary user-generated evaluations of businesses, products, or services, are a crucial form of customer feedback (Dixon, 2024). They significantly shape consumer behavior and influence purchase intentions (Ho & Dempsey, 2010). The importance of this topic is underscored by the multitude of studies seeking to understand the variables that could alter consumer behavior. These variables include review valence, review volume, review rating scores, perceived quality and reliability, and helpfulness votes, among others.

Homburg et al. (2015) focused on the role of review valence (positivity or negativity) and review volume. As part of their main findings, they concluded that both variables are significant, with a higher volume of positive (negative) reviews amplifying their positive (negative) impact. Furthermore, they highlighted the moderation role that product type plays, finding stronger effects on purchase intention from online reviews when users intending to buy experience goods compared to search goods. Following the importance of review valence, Schoenmueller et al. (2020) highlighted the polarized nature of these reviews. They concluded that the polarity is associated with the fact that consumers with strong opinions are more likely to write reviews.

Within the same line of findings, Filieri et al. (2021) highlighted that approximately 90% of online consumers consult online reviews before making purchasing decisions. They emphasized the importance of polarity ratings (extremely positive vs. extremely negative), as well as the effect on helpfulness votes. The authors point out that helpful reviews, which earn 'helpful votes' from other users, are crucial as they directly impact consumer purchase intentions and sales.

Xu et al. (2022) studied online reviews from both content creation and content consumption point of view. First, they learned how review characteristics and perceived helpfulness affect consumer involvement and engagement. Then, moving towards the creation side, they consider the role of reviewer expertise in determining review helpfulness.

These studies, along with others that will be discussed later, emphasize the importance and relevance of examining online reviews. The consistent findings across different research papers support the necessity of delving into this topic to understand dynamics that might not have been found yet, as well as different approaches to evaluating consistency. Given the proven impact electronic word of mouth has on marketing strategy, this paper focuses only on online reviews.

2.2 Anonymity vs. Identity disclosure

Anonymity has been defined as the inability of a group member to identify the origin (destination) of a message they received (sent). It is often considered as a binary or dichotomous factor, either being present or not. There are two types of anonymity: by process or content. The first one refers to the extent to which a person can determine the participation of another, while the second refers to the extent to which a person can identify the source of a contribution (Pinsonneault & Heppel, 1997).

Within the social media context, anonymity has been defined as a state where individuals share content without revealing their identities, allowing users to post without any identifiable detail associated with the content (K. Zhang et al., 2014). Data privacy and freedom of expression have been found to be the most important reasons behind a user's decision to post anonymously or reveal their identity (Pan et al., 2023). One main finding from K. Zhang et al. (2014) relates to anonymity leading to increased aggression and offensive content, as individuals are less constrained by social norms. However, Pan et al. (2023) found a positive social effect of anonymity in online content: they concluded that anonymity fosters an environment where individuals feel free to express themselves truthfully in moral or controversial contexts.

In the context of online reviews, identity disclosure has been widely studied as it has been proven to shape perceived credibility and trustworthiness, as it influences how the way community members judge reviews (see, e.g., Forman et al., 2008; W. Chen et al., 2019). Identity disclosure has been found to impact purchase intention through the sense of trust and authenticity generated (Homburg et al., 2015).

Reviewers' expertise has also been studied close to identity disclosure as some authors (see, e.g., Filieri et al., 2020; Xu et al., 2022) would argue there is no proven expertise without a full identity disclosure. Xu et al. (2022) emphasized the enhancement of perceived helpfulness of reviews as a combination of both identity disclosure and expertise, increasing consumer engagement. Filieri et al. (2020) assessed whether reviewer identity disclosure and reviewer expertise, as well as other factors, moderate the way users perceive helpfulness from extremely negative rating reviews.

2.3 Anonymity when producing content

There has been extensive literature that aims to study the effect of information disclosure on user-generated content in multiple situations, from social media profiles and posts to online reviews. However, as previously stated, this research will cover the effect of online reviews. Within the variety of perspectives that have been contemplated, the most studied factors are star rating, the sentiment perceived within the review's text, the quality or effort put into the review, how anonymity shapes the helpfulness of reviews, and the effect of previous anonymity on the decision of revealing the identity or not. This literature review examines the impact of anonymity on various aspects of online reviews, focusing on rating scores, sentiment scores, and review length. Table 1 presents a summary of the literature review examined.

Table 1: Literature Review - Anonymity when producing content.

Author/s Year Journal	Research Focus	Theoretical Framework	Sample / Method	Main Findings
Deng, L., Sun, W., Xu, D., & Ye, Q. (2021) Psychology & Marketing	The impact of an "anonymous review" feature on overall ratings, emotional expressions, and user behavior. It also explores how accumulated anonymous reviews affect subsequent user contributions.	- Deindividuation - Social Presence - Negativity bias	n = 19,860 reviews Restaurant review platform in China. - OLS - Fixed-Effect Panel Data	- Anonymity decreases ratings and positive emotions. - Accumulated anonymous reviews, have an impact on decreasing following ratings and emotions.
Hoyer, B., & van Straaten, D. (2022) Journal of Behavioral and Experimental Economics	The effect of anonymity versus pseudonymity on the number of ratings and how anonymity impacts the amount of market information and consumer welfare.	- Signaling theory - Intrinsic motives (altruism)	n = 192 subjects - ANOVA - Mann-Withney-U - $\chi 2$-test	- Anonymity is associated with a lower number of ratings. - Altruistic subjects are not affected by the introduction of anonymity. - Self-expression is blunted by anonymity.
Huang, N., Hong, Y., & Burtch, G. (2016) Fox School of Business Research	The effect of social network integration on user content generation: volume, exhibition of affective language and negative language.	- Social presence theory	n = 139,239 reviews Yelp.com and TripAdvisor.com - Difference-in-differences - Linguistic Inquiry and Word Count (LIWC)	- Social network integration increases the production of UGC and positive emotion in review text. - The integration reduces cognitive language, negative emotion, and expressions of disagreement (negations) in review text.
Pu, J., Chen, Y., Qiu, L., & Cheng, H. K. (2020) Information Systems Research	The effect of identity disclosure in content volume and average content length, as well as any possible cross-section effects.	- Social presence theory - Inhibition effect - Displacement effect	n = 591 users 36,107 reviews 429,857 answers - Regression models with fixed effects - Panel Data - Time Series	- Disclosing user identities increased the effort users put into each piece of content, while decrease the number of reviews.

Rating scores in online reviews are considered the fastest cue of the overall customer experience regarding a product, service, or business. Many e-commerce sites allow users to leave a star rating of one to five (Matsakis, 2019). Although the overall product rating score is usually calculated using an advanced machine learning process, as is the case for Amazon, its helpfulness is undoubted. They influence consumer decision-making, as potential buyers often rely on these scores to assess the quality and reliability of a product or service before making a purchase.

Online ratings are a crucial aspect of online reviews. They play a vital role in helping consumers reduce uncertainty and assess the risks associated with unfamiliar products and services, shaping price sensitivity, risk perception, and even trust (Zimmermann et al., 2018). Given their importance, the impact of anonymity on rating scores is a critical area of study, as it directly influences the perceived credibility and usefulness of reviews. Researchers have extensively examined how varying levels of identity disclosure affect the ratings users assign to products and services.

Deng et al. (2021) studied the effect of anonymity on online reviews through the introduction of an "anonymous review" function on a platform. Using restaurant-specific data and drawing on social presence and negativity bias theories, they concluded that implementing the anonymity function leads to a decrease in overall ratings. Likewise, the effect was found to increase over time as subsequent reviewers, influenced by negative reviews, were more likely to provide lower ratings themselves. Another way this effect has been studied is by increasing the level of information disclosure through social network integration (Huang et al., 2016). This integration led to more reviews, which, due to the increased social presence, raised rating scores overall.

Furthermore, other authors have found that the rating score itself is highly correlated with the review sentiment, suggesting that short reviews with more positive sentiments tend to lead to higher rating scores (Ghasemaghaei et al., 2018). Other studies have evaluated the effect of anonymity across different business models. Gutt and Neumann (2019) analyzed this effect within the B2B model. The authors analyzed online reviews on a platform for B2B where users could decide to fully reveal their identity, as well as the company information, or hide everything. Aligned with findings by Deng et al. (2021), they concluded that anonymity was associated with lower ratings and the effect held across different industries. Teubner & Glaser (2018) analyzed review data from Airbnb, strengthening the previous findings of a higher rating score in the presence of identity disclosure. Overall, it has been found a negative correlation, between information disclosure and rating scores. Thus, the following hypothesis is proposed:

H1: *Anonymous reviewers tend to give lower rating score compared to other reviewers.*

In addition to the change in overall ratings associated with the anonymity effect, as Ghasemaghaei et al. (2018) concluded, there is an interesting relationship between rating scores and sentiment analysis. Sentiment score is often defined as a numerical assessment of the emotional tone. It is calculated through supervised or unsupervised methods, applied to the content text following, in most cases, a polarity approach (Ashtiani & Raahemi, 2023).

Sentiment analysis on a text could be measured through different methodologies, from polarity classification (positive, neutral, and negative) to sentiment scores, both methodologies supported by linguistic analysis. East et al. (2007) found a strong positivity bias on review sentiment when analyzing extremely positive or negative reviews. The findings highlight the predominance of positive over negative WOM and the association between market share and WOM incidence: if a category is often discussed negatively, then it is also likely to be addressed positively.

Filieri et al. (2018) shaped their research into the moderating factors shaping extremely negative online reviews and perceived helpfulness. They found that aligned with what has already been seen, identity disclosure, as well as reviewer expertise, readability, and review length, increases the perceived helpfulness of negative reviews by affecting the sentiment expressed. Furthermore, as has already been presented, Hoyer and Van Straaten's (2022) study finds that anonymity significantly reduces the number of ratings provided by users and affects the sentiment expressed in reviews. Thus, the following hypothesis is proposed:

H2: *When the reviewer is perceived as non-anonymous, a social desirability effect influences the review content to have a more positive sentiment than reviews where the reviewer is anonymous.*

Review length, often measured by the number of characters or words, is another interesting indicator of the effort the reviewer put into it. Longer reviews typically give more in-depth information, reducing the information asymmetry present in online markets. Review length can influence the way potential buyers perceive the helpfulness of the review while also influencing the purchase decision.

Ghasemaghaei et al. (2018) highlighted the fundamental role of review length in consumer decision-making processes, particularly finding that longer reviews are often more critical in reducing customers' searching costs. Their analysis was given in the insurance sector. Also drawing attention to the importance of length, Zhang et al. (2019) concluded that reviewing content richness, measured through length and pictures, influences booking decisions.

Pu et al. (2020) studied the effects of identity disclosure on content generation in an online community, measuring the effort of a review through its length. They concluded that identity disclosure created a displacement effect, increasing effort but decreasing the content volume. Supporting these results, Zhang et al. (2023) concluded that longer reviews are positively correlated with reviewers' preference for anonymity. Thus, the following hypothesis is proposed:

H3: *Reviewer's anonymity leads to longer reviewer text.*

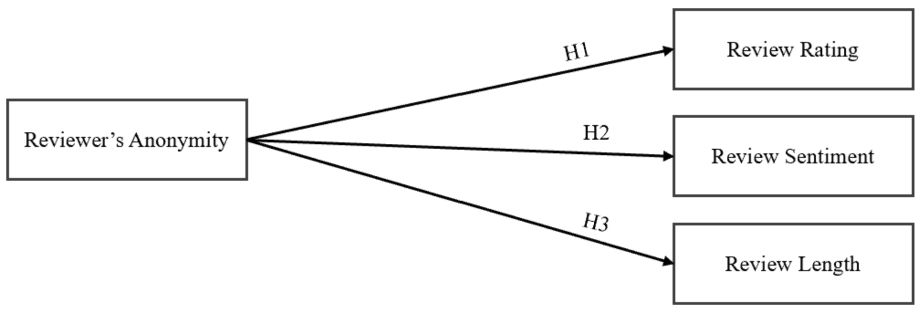

Figure 1: Conceptual Framework - Anonymity when producing content.

2.4 Anonymity when consuming content

In addition to its impact on content creation, anonymity also plays a significant role in how users consume information on online platforms. This section of the literature review explores how the anonymity of reviewers and the presence of anonymous content influence other users' engagement, trust, and consumption behavior. Specifically, how perceived reviewer reliability mediates the impact of anonymity on purchase intention and how review sentiment moderates this relationship. It also examines the moderating role of reviewing sentiment on the relationship between anonymity and purchase intention. Table 2 presents a summary of the main existing literature.

Purchase intention refers to the level to which a consumer is likely to buy a product in a physical or online store. It is a key factor when analyzing actual purchase behavior, analyzing the different elements that play a role in the pre-purchase stage (Peña-García et al., 2020). Within the context of online marketplaces, Rosario et al. (2016) highlighted the importance of factors such as review valence (positive or negative), volume, and helpfulness on purchase intention. Their findings concluded that electronic word-of-mouth volume has a stronger effect on sales than valence. Similarly, Vana and Lambrecht (2021) found a significant influence of review star ratings and their position on product pages on purchase intention. Particularly, they found that top-positioned reviews, with high star ratings, significantly increase purchase likelihood.

Table 2: Literature Review - Anonymity when consuming content

Author/s Year Journal	Research Focus	Theoretical Framework	Sample	Main Findings
East R., Hammond K. & Lomax W. (2008) Intern. J. of Research in Marketing	The effects of positive and negative WOM on brand evaluations and purchase probability.	- Negativity effect	n = 2,544 respondents - Wilcoxon Tests - Regression Analysis	- Positive WOM usually had more impact than negative WOM on purchase intention
Forman, C., Ghose, A., & Wiesenfeld, B. (2008) Information Systems Research	Effects of identity disclosure on manner, structure of the reviewer on the helpfulness rating given to the review.	- Social identity theory	n = 175,714 Amazon reviews - 2SLS	- Information disclosure enhances the perceived helpfulness of reviews and positively affects sales.
Hong, C., & Li, C. (2017) Journal of Language and Social Psychology	Differences in trust, processing depth, and behavioral intention when the message is anonymous vs. when the user can be identified.	- Language expectancy theory - Information processing theory - Attribution Theory	n = 161 Between subjects experiment -ANCOVA	- An anonymous message generates more trust. - Positive messages increase behavioral intention more effectively than negative messages.
Kusumasondjaja S., Shanka T. & Marchegiani C. (2012) Journal of vacation marketing	Effect of valence of online reviews (positive vs. negative) and the presence of reviewer identity information, influence consumer perceptions of review credibility and initial trust in travel services.	- Social Identity Theory - Valence Framing Theory - Elaboration Likelihood Model	n= 639 travelers Between subjects experiment - ANOVA	- Negative reviews are found to be more credible, while positive reviews lead to greater initial trust, but this only applies when the reviewer's identity is disclosed.

Another factor when analyzing online reviews' effectiveness is *perceived reliability* and its role in purchase intention. Trust has been highlighted as a key factor in the seller-buyer relationship of any transaction type. The anonymity and lack of in-person contact in the online marketplace have increased how potential buyers value trust in the review (Pavlou & Dimoka, 2006). An interesting point presented by Martin (2017) states that the reliability perceived by the receiver will have a direct effect on the way the receiver behaves, encouraging them to act consistently with the recommendations offered.

Pavlou and Dimoka (2006) conducted a comprehensive analysis of the feedback process, revealing how it shapes price premiums through the mediating role of trust. Their findings significantly bolster the model that links feedback with price premiums, demonstrating that text comments hold substantial influence over buyers' trust in sellers' credibility. Similarly, Martin (2017) delved into the influence of positive and negative WOM on user perceptions and customer decisions, finding a positive bias where positive WOM carries more weight in shaping consumer attitudes, especially when it comes from identifiable reviewers, who are perceived as more reliable.

Exploring this effect in greater detail, Forman et al. (2008), in an online market context, analyzed the effect of disclosing personal information on helpfulness ratings and subsequent product sales. Using extensive data from Amazon, they concluded that anonymity might reduce perceived reliability and, consequently, purchase intention. These results are also aligned with Pavlou and Dimoka (2006), who highlighted the importance of feedback text comments in building trust in online marketplaces. Thus, the following hypothesis is proposed:

H4: *Perceived reliability mediates the impact of anonymity on purchase intention.*

Review sentiment is another critical factor that influences perceived reviewer reliability. Fan et al. (2021), in their study of attribution theory through online booking platform data (TripAdvisor), concluded that reviews with extreme sentiments (positive or negative) are perceived as more helpful than those in between or two-sided. Furthermore, Filieri et al. (2018) concluded that the presence of reviewer identity moderates the perceived helpfulness of extremely negative reviews.

Martin (2017) concluded that positive word-of-mouth increases the reliability of the reviewer when the reviewer discloses information. This suggests that positive reviews from identifiable sources are perceived as more reliable. Contrarily, Kusumasondjaja et al. (2012) analyzed how reviewers' identity information could affect perceived credibility, concluding that when the reviewer's identity is not disclosed, there is no significant difference between positive and negative reviews in terms of perceived credibility. However, they concluded that, in general, negative reviews are perceived as more credible. Thus, the following hypothesis is proposed:

H5: *Review sentiment moderates the relationship between reviewer anonymity and perceived reliability. Specifically, when a review is more positive, consumers perceive higher reliability.*

Review sentiment is considered a key factor not only for shaping the reliability perceived by users but also for the final purchase intention. East et al. (2008) analyzed the direct effect of positive and negative WOM on brand purchase likelihood. They concluded that positive WOM usually had more impact than negative WOM on purchase intention. Furthermore, Lei et al. (2023) investigated the effect of exposure to positive or negative reviews on product sales. They found that users have a negative bias in the information-seeking face. However, users consider positive reviews to be more helpful when there is a confirmation bias.

Moreover, Zhang et al. (2019) investigated the effect of online reviews on the timing of restaurant bookings. They concluded that higher rating variation and review content lead to earlier bookings. Thus, the following hypothesis is proposed:

H6: *Review sentiment moderates the relationship between anonymity and purchase intention.*

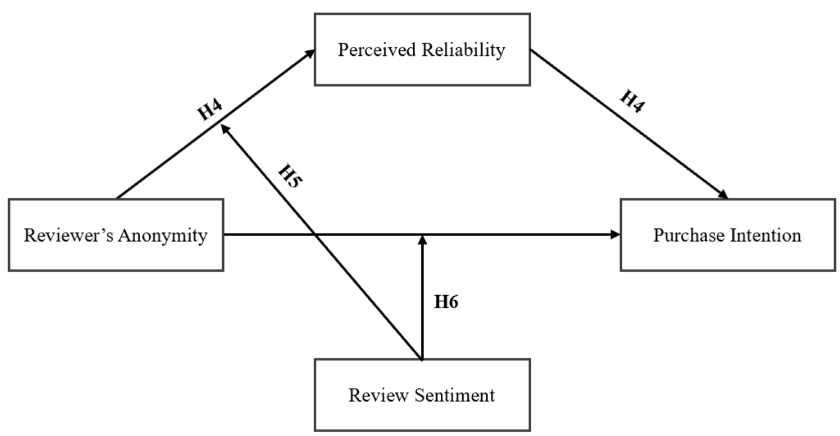

Figure 2: Conceptual Framework - Anonymity when consuming content.

3. Study 1: Anonymity when producing content (Secondary Data)

3.1 Data

The first part of this research involves secondary data from Amazon.com, initially published by Ni et al. (2019). The dataset contains online reviews from May 1996 to October 2018 and consists of 233.1 million reviews for 29 different product categories. For the objective of this research, multiple data transformations were implemented. The data structure consisted of individual compressed JSON files by category. Due to the size of the files, the initial data transformation was performed in Python using the library PySpark.

Initially, to keep more recent data, only reviews provided since 2008 onwards were considered, and duplicates (identified based on reviewer ID, product category, rating score, review text, and review date/time) were removed. Then, a threshold was set at 500,000 reviews per category. To achieve this, two different processes were performed: First, for product categories with less than 500,000 reviews, all observations that met the first condition were kept. The second step involved setting a threshold for a random sample of 500,000. However, due to the nature of the PySpark library used for handling large datasets, it did not directly provide an exact 500,000 random sample. Instead, it calculated this desired sample as a proportion of the total dataset and then selects a random sample based on that percentage. As a result, the sizes of the data samples across different categories varied slightly, but they consistently floated around the 500,000 mark.

This process resulted in a total sample of 13.3 million reviews across 29 product categories. As set by Amazon, reviews extend a rating score range of 1 to 5, with an average rating score of 4.2. The average review length stands at 247 characters, reflecting detailed consumer feedback on most of the cases. Additionally, it's interesting to notice that 99.9% of the total reviews considered in the sample contained review text. This high proportion of reviews containing textual content could indicate that the dataset is well-suited for the current study. Furthermore, the dataset encompasses a vast array of contributors, with 8.0 million unique reviewers' identification codes (reviewerID) and 4.4 million distinct usernames, underlining the diversity and breadth of the database. Table 3 provides a summary of the measures mentioned before by product category.

In line with prior research examining the impact of perceived anonymity, a username classification is essential to categorize reviewers' names as either anonymous or authentic[1]. Previous studies have employed various methods, including leveraging natural experiments prompted by new technological implementations (Deng et al., 2021), defining anonymous usernames subsequently used in experimental designs (Hoyer & Van Straaten, 2022), or conducting classification processes through rule-based methods executed by research assistants

[1] Authentic is used in this context to refer to names that appear genuine or real according to the defined criteria.

(Jiang et al., 2022). In this study, classification criteria were established using the 'gender()' package in R along with additional rule-based methods.

Table 3: Summary Statistics by product category

Product Category	# Unique ReviewerID	# Reviews	Avg Rating	Avg Review Length
Amazon_Fashion	450,148	489,701	3.90	148.08
All_Beauty	324,038	346,712	4.09	207.55
Appliances	437,168	485,640	4.27	184.68
Arts_Crafts_and_Sewing	401,579	488,931	4.31	155.90
Automotive	439,590	497,750	4.25	168.63
Books	438,373	499,477	4.39	455.38
CDs_and_Vinyl	358,151	494,225	4.49	512.67
Cell_Phones_and_Accessories	475,291	497,862	3.93	201.42
Clothing_Shoes_and_Jewelry	472,862	496,854	4.19	158.11
Digital_Music	347,003	478,238	4.65	219.12
Electronics	468,986	497,760	4.08	286.28
Gift_Cards	128,877	143,588	4.67	98.58
Grocery_and_Gourmet_Food	427,949	495,188	4.31	175.48
Home_and_Kitchen	469,899	498,129	4.20	202.64
Industrial_and_Scientific	432,053	491,887	4.29	185.54
Kindle_Store	355,153	498,222	4.30	386.21
Luxury_Beauty	372,724	454,208	4.22	219.94
Magazine_Subscriptions	72,098	85,292	4.03	247.07
Movies_and_TV	412,526	494,393	4.23	322.90
Musical_Instruments	379,653	488,739	4.25	270.18
Office_Products	455,480	495,973	4.17	203.14
Patio_Lawn_and_Garden	451,098	496,862	4.12	207.88
Pet_Supplies	431,725	496,378	4.15	231.98
Prime_Pantry	247,659	427,750	4.32	138.07
Software	375,147	446,715	3.56	430.23
Sports_and_Outdoors	464,290	496,787	4.24	209.58
Tools_and_Home_Improvement	453,871	497,287	4.22	214.14
Total	7,978,523	13,270,091	4.20	246.82

For general context, the 'gender()' package aims to categorize names by gender (female or male) based on historical data collected on U.S. Census data from the Minnesota Population Center at the University of Minnesota (Mullen, 2021), despite the multiple limitations (see, e.g., Kozlowski et al., 2022; Lockhart et al., 2023; Mohammad, 2020) of this package, the primary purpose of this package in this research was to determine if the reviewer's name was found in the gender database. In this context, the assumption is that if a name appears in this database, it exists, suggesting it's not an anonymous name and is categorized as authentic. Despite the literal meaning of the word "authentic" in the context of this research, a name is considered authentic if, based on the logic and rules defined, it is not regarded as anonymous. Furthermore, the possibility that a user might select a username that appears authentic but differs from their real name falls outside the scope of this research.

For the first criterion, associated with the 'gender()' package, there was a need to extract the names from the reviewer's name field (reviewerName). This process was done using three rules that were sequentially applied:

1. All numbers and special characters were removed (Step 1).
2. The first word preceding a space was selected (Step 2).
3. From this resultant word, only the first word preceding a capital letter was retained in case there was any (Step 3).

For example, the reviewerName "LauRence Sutton25", taken from the dataset, turns to "LauRence Sutton" after Step 1, to "LauRence" after Step 2, and finally, "Lau" after Step 3. These rules were designed to extract the most accurate name from the reviewerName field, which could then be processed using the 'gender()' package: if the name defined corresponded to a name with a gender association in the 'gender()' package, it was classified as authentic.

The second criterion, based on rule-based criteria, assumed a name was verified if it started with a capital letter followed by a period, a space, and at least three additional letters. This rule applied to cases where users opted to hide their first name but retained an authentic name structure.

The methodology for username classification was put to the test with one hundred randomly selected names. These names were sent to two individuals unrelated to this research, who were asked to classify them as either Anonymous or Authentic. The level of agreement between the different classifications was then analyzed using Cohen's Kappa measure. Overall[2], the classification suggested an either substantial agreement (between 0.61 to 0.80) or perfect agreement (> 0.80) (McHugh, 2012).

[2] Cohen's Kappa between the two participants was 0.900. Between one participant and the global methodology, it was 0.817, and with the other participant, it was 0.798.

Overall, 62.8% of reviewer names were classified as authentic, while 37.2% were anonymous; correspondingly, 41.0% of total reviews are associated with an anonymous reviewer. In fact, anonymous reviewers wrote, on average, slightly more reviews than authentic reviewers (0.11 more than authentic reviewers). It's important to keep in mind that a single reviewer ID may have different reviewer names for other reviews, as reviewers can change their public names whenever they want without restrictions. Table 4 presents detailed statistics by category.

Table 4: Statistics by Username Category

Username Category	# Unique ReviewerID	# Unique ReviewerName	# Reviews	Avg Rating	Avg Review Length	Avg #Reviews
Anonymous	3,151,088	1,634,231	5,443,703	4.18	265	1.73
Authentic	4,839,780	2,757,811	7,826,388	4.22	234	1.62

On average, reviews written by anonymous users are 31 characters longer than authentic ones. This difference becomes more pronounced when considering ratings. Anonymous reviewers consistently provide longer feedback across all rating categories, with particularly substantial reviews for lower ratings. For instance, anonymous reviews for rating 1 are, on average, 86 characters longer than those for rating 5. Furthermore, anonymous reviews for rating 5 are significantly shorter than those for other ratings, averaging only 224 characters, while the lengthiest anonymous reviews are associated with rating 2, averaging 352 characters. In contrast, authentic reviews exhibit a symmetrical frequency behavior, showing relatively shorter reviews for both ratings 1 and 5, followed by the highest length for ratings 2 and 4 (307 characters each), and finally 304 characters for rating 3. Table 5 illustrates these findings.

Table 5: Avg Review length (by characters) by category and review rating score (1 to 5)

	Rating Score				
	1	**2**	**3**	**4**	**5**
Anonymous	309	352	347	342	224
Authentic	265	307	304	307	201

The general distribution of anonymous versus authentic usernames remains consistent across the different rating options: on average, 41.9% of reviews were written by reviewers categorized as anonymous. Figure 3 illustrates the total number of reviews, in millions, by rating and category.

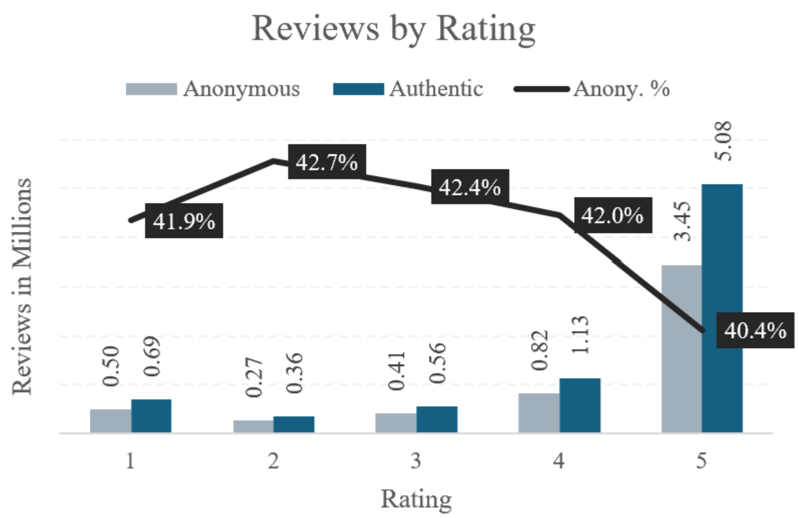

Figure 3: Reviews in millions by rating score and percentage of anonymous.

As previously found (Filieri et al., 2018; Huang et al., 2016; Schoenmueller et al., 2020; Vana & Lambrecht, 2021; Zhang et al., 2023), people generally tend to leave higher ratings, and this dataset is no exception, with 64.3% of total reviews being rated with five stars. When analyzing the distribution by category, a small difference emerges: 63.3% of reviews written by anonymous reviewers were allocated to the 5-star ranking, compared to 64.9% when authentic reviewers wrote the reviews.

On Amazon, users have the option to customize their public name, which is associated with contributions such as Customer Reviews. This feature allows users to choose whether to display their real name or opt for any other pseudonym. Consequently, it implies that multiple users can share the same Public Name, offering flexibility and privacy to individuals while engaging with the platform (Can I Review Anonymously at Amazon ?, 2016; About Public Names - Amazon Customer Service, n.d.).

When analyzing the top 5 review names for each category, it becomes evident that within the anonymous names, a notable proportion of reviews were attributed to *Amazon Customer*, with 18.5%. However, the subsequent name *Kindle Customer*[3] accounted for only 1.8%, indicating that

[3] Kindle Customer is a username available in all product categories, not only under the Kindle category.

despite a strong preference for the *Amazon Customer* username, anonymous reviews are distributed among various alternative names. It's important to consider that Amazon sets a default name, *Amazon Customer*, when a customer has not set a profile name. This could explain the high frequency of reviews written by that anonymous name.

Moreover, among authentic names, the top 5 collectively represent only 0.8% of the total reviews written by users with names classified as authentic. This suggests a widespread distribution of data and a low frequency of authentic names, emphasizing the diversity and inclusivity of user identities within the dataset.

Table 6: Top 5 ReviewNames by category

Anonymous		**Authentic**	
ReviewerName	**# Reviews**	**ReviewerName**	**# Reviews**
Amazon Customer	1,009,375	Mike	13,186
Kindle Customer	99,944	Chris	13,053
Pen Name	6,152	John	12,899
Anonymous	5,664	David	11,082
J	4,770	Michael	10,393

In addition to the previously mentioned data, the Amazon dataset also includes a verification field categorized as 'true' or 'false,' along with vote counts. The 'verified purchase' label indicates whether the reviewer has bought or used the item on Amazon and paid the price available to most Amazon shoppers (Amazon Verified Purchase Reviews, n.d.). Within the database, 85.2% of reviews are categorized as verified purchases. When validating the values for anonymity and authenticity, the former has a share of 84.3% of verified reviews, while the latter consists of 85.9%. A chi-square test reveals a statistically significant difference between these two categories (p-value < 0.001). Furthermore, logistic regression analysis indicates that reviewers classified as 'Authentic' are less likely to be verified buyers or users on Amazon compared to those classified as 'Anonymous' (p-value < 0.001).

The variable 'vote' represents the number of users who found the review helpful by clicking a button located below the review. This option is not mandatory, and reviews may not receive any helpful votes. In fact, only 14% of the reviews in the database received votes, totaling 11.4 million reviews with at least one helpfulness vote. The distribution remains consistent between anonymous and authentic reviews, with 15% for the former and 14% for the latter. Furthermore, the t-test indicates a statistically significant difference (p-value < 0.001) between anonymous and authentic. Among the reviews with votes, the average was seven votes, with a minimum of 2 votes and a maximum of 999 votes. This resulted in a standard deviation of 20 votes.

Understanding the factors that contribute to a review's helpfulness is crucial, and sentiment analysis provides additional insights into the subjective content of the reviews. Sentiment analysis is a natural language processing technique that aims to automatically identify, extract, and quantify subjective information, such as opinions, emotions, and attitudes, expressed in text data. Supervised sentiment analysis relies on labeled data to train machine learning models to predict sentiment, while unsupervised methods use algorithms to uncover patterns and sentiments within text without labeled training data. As Diamantini et al. (2019) concluded, unsupervised methods are beneficial for sentiment analysis of social content as they are dynamic and cover a wide range of topics. Therefore, sentiment scores were computed using the Natural Language Toolkit (NLTK) in Python, following the polarity method fitting an unsupervised method. This method assigns sentiment scores on a scale ranging from -1 to 1, where negative scores denote negative sentiment, positive scores indicate positive sentiment, and scores around zero represent neutral sentiment (NLTK:: Release Notes, 2023).

The initial sentiment analysis was conducted using the review text from the 'reviewText' column of the dataset. Consistent with existing literature on positivity bias (East et al., 2007; Schoenmueller et al., 2020), the average sentiment score when creating content is 0.502 across all categories, which suggests a positive bias across the dataset. Moreover, 10.6 million reviews have a sentiment score larger than 0 (positive score), compared to 1.6 million reviews with negative scores, followed by 1.0 million reviews with neutral scores.

When analyzed by rating, as shown in Table 7, it's clear that there is consistency in the average sentiment score and the rating category. The average sentiment score does not seem to differ between anonymous and authentic reviewers. However, when analyzing the difference by status and rating, some ratings exhibit a higher difference than others.

Table 7: Sentiment Scores by Rating Score (1 to 5) and Reviewer Category (Anonymous vs. Authentic)

Category	1	2	3	4	5	Overall
Anonymous	-0.107	0.080	0.294	0.552	0.637	0.503
Authentic	-0.117	0.061	0.282	0.544	0.632	0.502
Overall	-0.113	0.069	0.287	0.548	0.634	0.502

Table 8 provides a comprehensive summary of the main dependent variables (DV) analyzed in this study, encompassing rating scores, sentiment scores, and review lengths. This reflects the diverse range and distribution of data collected and discussed in the preceding sections.

Table 8: Summary descriptive statistics for DV variables

DV	Min	Mean	Median	Max	Q25	Q50	Q75
Rating Score	1	4.20	5	5	4	5	5
Sentiment Score	-1	0.50	0.64	1	0.28	0.64	0.86
Review Length	0	246.65	119	33,915	44	119	260

3.2 Results

In this study, we analyzed and examined the differences between anonymous and authentic content generation, specifically focusing on online reviews. The aim was to investigate the effect of anonymity on the nature of user-generated content. Our findings revealed significant distinctions between anonymous and authentic reviewers, indicating that anonymity does indeed influence the content generated. Furthermore, to delve deeper into the magnitude of this effect, hypothesis testing was conducted using different methodologies. This allowed us to quantify the size of the effect and gain a more comprehensive understanding of how anonymity impacts the generation of online reviews.

Fixed effects estimation involves accounting for unobserved individual-level differences by including individual-specific dummy variables, thereby capturing the effects of variables that are constant within individuals over time, such as time-invariant characteristics or unmeasured factors (Encyclopedia of Social Measurement, 2005). As previously stated by many authors studying online reviews and UGC in general, the most common fixed effect in this context is time-based (Anderson & Simester, 2014; Deng et al., 2021; Homburg et al.,2015; Huang et al., 2016; Wang & Chaudhry, 2018; Zhang et al., 2023). Therefore, due to the composition of the data, a fixed effect controlling possible seasonality effects based on month is added. Furthermore, another popular fixed effect is category or product type (Anderson & Simester, 2014; Forman et al., 2008; Vana & Lambrecht, 2021). Since the dataset of this research contains 29 product categories, the second fixed effect considered is differences based on categories.

As previously reported, rating scores did not have a normal distribution, which did not allow for running an ANOVA test. Therefore, a Kruskal-Wallis test was implemented (Hoyer & Van Straaten, 2022), revealing significant differences between groups (chi-square = 2598.46 and p-value <0.001). To understand in depth the effects, due to the lack of normality on the dependent variable and its nature of a categorical variable where order matters, an ordinal logistic regression was implemented. With an overall p-value of 0.000 for all models, considering an alpha of 0.05, all regressions are considered significant. Based on the ordinal logistic regression analysis, there are significant effects of both authenticity and verification of the purchase on the rating scores. Specifically, the authenticity of the reviewer's name (classified as authentic vs. anonymous) has a small but statistically significant positive effect on the rating score (on average, a coefficient of 0.063 with p-value < 0.001). This suggests that reviews from authentic-seeming names are likely to have slightly higher ratings compared to anonymous ones, as other authors have found (see, e.g., Deng et al., 2021; Gutt & Neumann, 2019; Teubner and Glaser, 2018). Furthermore, whether the purchase was verified through Amazon shows a more substantial positive impact on the rating score (on average, a coefficient = 0.347, p-value < 0.001), implying that verified purchases are associated with significantly higher ratings.

Table 9 presents a summary of the results for Equation 1 following 4 different variations: (A) without considering fixed effects, (B) considering only monthly fixed effects, (C) considering only category fixed effects and (D) both fixed effects included.

Equation 1: Rating Score complete model

$$Rating_i = \beta_0 + \beta_1 Authentic_i + \beta_2 Verified_i + \sum_{j=1}^{28} \beta_{j+2} Category_i + \sum_{k=1}^{11} \beta_{k+30} Month_i + \varepsilon_i$$

With an overall p-value of 0.000 for all models, considering an alpha of 0.05, all regressions are considered significant. Based on the ordinal logistic regression analysis, there are significant effects of both authenticity and verification of the purchase on the rating scores. Specifically, the authenticity of the reviewer's name (classified as authentic vs. anonymous) has a small but statistically significant positive effect on the rating score (on average, a coefficient of 0.063 with p-value < 0.001). This suggests that reviews from authentic-seeming names are likely to have slightly higher ratings compared to anonymous ones, as other authors have found (see, e.g., Deng et al., 2021; Gutt & Neumann, 2019; Teubner and Glaser, 2018). Furthermore, whether the purchase was verified through Amazon shows a more substantial positive impact on the rating score (on average, a coefficient = 0.347, p-value < 0.001), implying that verified purchases are associated with significantly higher ratings.

Table 9: Rating Scores Analysis

DV = Rating Score						
Model	(A)	(B)	(C)	(D)	(E)	(F)
Method	Ordinal Logistic Regression				Logistic Reg.	Linear Reg.
Authentic (binary)	0.063*** (0.001)	0.062*** (0.001)	0.064*** (0.001)	0.064*** (0.001)	0.011*** (0.000)	0.033*** (0.001)
Verified	0.321*** (0.002)	0.320*** (0.002)	0.373*** (0.002)	0.372*** (0.002)	0.056*** (0.000)	0.249*** (0.001)
Fixed Effect (Month)		Included		Included	Included	Included
Fixed Effect (Category)			Included	Included	Included	Included
Observations	13,270,091	13,270,091	13,270,091	13,270,091	13,270,091	13,270,091
LR chi2(41)	47,385	47,959	358,653	359,275		
Prob > chi2	0.000	0.000	0.000	0.000	0.000	0.000

Note: ***p<0.01, **p<0.05, *p<0.1

For robustness, two additional analyses were performed. First, a new binary variable was created, coded as 1 when the rating score was 5 and 0 otherwise. The coefficients for both authenticity and purchase verification remained positive and statistically significant (See Table 9 Model E for details of the results). Specifically, an authentic-seeming reviewer's name increases the likelihood of receiving a 5-star rating (coefficient = 0.068, p-value < 0.001), consistent with the ordinal regression finding that authenticity positively impacts rating scores. Similarly, a verified purchase significantly increases the likelihood of a 5-star rating (coefficient = 0.351, p-value < 0.001), reinforcing the earlier result that verification has a strong positive effect on ratings. These results confirm the robustness of the original findings, emphasizing the influence of both authenticity and verification on higher ratings.

As a second robustness check, the widely accepted statistical method of ordinary least squares (OLS) regressions was employed despite the lack of normality, as other authors have done (see, e.g., Deng et al., 2021; Huang et al., 2016; P.-Y. Chen et al., 2018). This approach was used with both raw and logarithmic transformations on the dependent variable. Once more, the coefficients for both authenticity and purchase verification remained positive and statistically significant (p-value < 0.001). For OLS with raw data, the coefficient for authentic-seeming status was 0.033, and for verified purchase, it was 0.249 (See Table 9 model F for details of the results). Furthermore, when using a logarithmic transformation, the coefficients were 0.010 and 0.092, respectively.

A common factor found across many studies is the high correlation between rating score and sentiment score, meaning it is expected to have a high sentiment score when a high rating score was provided (see, e.g., Deng et al., 2021; Ghasemaghaei et al., 2018). The dataset used for this study corroborated this theory with a moderate correlation between sentiment scores and rating scores of 0.5107, indicating a significant positive relationship.

The sentiment score and rating score are not normally distributed. The positivity bias present in approximately 80% of the data does not allow for testing differences across groups with an ANOVA test. Furthermore, due to the nature of the variable (continuous between -1 and 1), a Wilcoxon rank-sum (Mann-Whitney) test was performed. The rank sums showed that the rank sum for anonymous reviews is slightly lower than expected, while the rank sum for authentic reviews is marginally higher than expected. This suggests that authentic reviews tend to have higher ranks (and thus higher sentiment scores) compared to anonymous reviews (p-value <0.0001). To get a better sense of the factors affecting this relationship and relying on the lack of normality, an analysis through robust regression was conducted for additional insights.

In summary, the Robust Regression results present a consistent and significant (p-value < 0.001) negative coefficient for authenticity across all models. This finding has significant implications for the field of sentiment analysis and consumer behavior, as it suggests a lower sentiment score for reviewers classified as authentic. Control variables verified and rating score also consistently show significance. Firstly, reviews marked as verified are found to be associated with a lower sentiment score, indicating more negative emotions presented in the review. A possible explanation for this effect is that reviewers with a verified purchase are more likely to write a review when they did not like the product, increasing the chance of more negative emotions, thus reducing the sentiment score. Finally, the rating score, as the correlation results suggested, has a positive significant (p-value <0.000) relation with the sentiment score, increasing the sentiment score as the rating score increases. These findings underscore the importance of considering authenticity, verification, and rating scores in sentiment analysis and consumer behavior studies. Table 10 presents results for Equation 2 following the same four variations implemented for Rating Scores.

Equation 2: Sentiment Score complete model

$$\begin{aligned} Sentiment\_Score_i &= \beta_0 + \beta_1 Authentic_i + \beta_2 Verified_i + \beta_3 RatingScore_i \\ &+ \sum_{j=1}^{28} \beta_{j+3} Category_i + \sum_{k=1}^{11} \beta_{k+31} Month_i + \varepsilon_i \end{aligned}$$

Robust regression results present a consistent and significant (p-value < 0.001) negative coefficient for authenticity across all models, suggesting a lower sentiment score for reviewers classified as authentic. Control variables, verified and rating score are also consistently significant. Firstly, reviews marked as verified are found to be associated with lower sentiment scores, indicating more negative emotions presented in the review. A possible explanation for this effect consists of reviewers with a verified purchase being more likely to write a review when they did not like the product, increasing the chance of more negative emotions and reducing sentiment scores. Finally, the rating score, as the correlation results suggested, has a positive significant (p-value <0.000) relation with the sentiment score, increasing the sentiment score as the rating score rises.

Furthermore, in the robustness checks, some modifications of the assumptions were made. One of these was the exclusion of neutral sentiment scores (sentiment score =0) for two key reasons. Firstly, as it has been previously studied, managerial and actionable implications are associated with either positive or negative perceived feelings (see, e.g., East et al., 2007; Filieri et al., 2018; Ghasemaghaei et al., 2018; Huang et al., 2016). Secondly, neutral reviews constitute a small proportion of the data set, with approximately 7.8% of total reviews, implying that the exclusion does not compromise the robustness of the study. Running a robust regression analysis with both fixed effects and the same independent variables but limiting to polar sentiment scores resulted in consistent conclusions (see Table 10 model E for details of the results).

Table 10: Sentiment Score Analysis

	DV: Sentiment Score				
Model	(A)	(B)	(C)	(D)	(E)
Method	Robust Regression				Robust Reg. no neutral sentiment
Authentic (binary)	-0.010***	-0.010***	-0.011***	-0.011***	-0.010***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Verified	-0.157***	-0.157***	-0.133***	-0.133***	-0.119***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Rating Score	0.200***	0.200***	0.199***	0.199***	0.205***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Fixed Effect (Month)		Included		Included	Included
Fixed Effect (Category)			Included	Included	Included
Observations	13,260,678	13,260,678	13,260,678	13,260,678	12,216,860
Prob > F	0.000	0.000	0.000	0.000	0.000

Note: ***p<0.01, **p<0.05, *p<0.1

The study's robustness was further validated by implementing a categorical variable, grouping rating scores as negative (sentiment score <0), positive (sentiment score >0), or neutral (sentiment score = 0). This new category allowed us to perform a multinomial logistic regression, with positive scores as the comparison group. The key findings were significant, indicating that reviews written by authentic reviewers are less likely to be negative (coefficient = -0.009, p-value <0.001) and more likely to be neutral (coefficient = 0.085, p-value <0.001) compared to positive reviews. This suggests that authentic reviewers tend to provide more balanced feedback, avoiding extreme negativity but sometimes opting for neutral over positive. On the other hand, verified reviews are more likely to be negative (coefficient = 0.107, p-value <0.001) and significantly more likely to be neutral (coefficient = 0.868, p-value <0.001) compared to positive reviews. This indicates that verified purchasers might have higher expectations and thus provide more critical feedback. Additionally, higher rating scores (closer to 5) are associated with a higher likelihood of the review being positive rather than neutral (coefficient = -0.488 , p-value <0.001) or negative (coefficient= -0.918 , p-value <0.001)[4].

Despite the initial expectations, based on the moderate correlation between sentiment score and rating score, that authentic reviewers would have higher sentiment scores, all robust regressions consistently showed a negative coefficient for authenticity. This leads to the conclusion that anonymous reviewers tend to write with overall more positive emotions, leading to higher sentiment scores compared to authentic reviewers. One possible explanation might be that authentic reviewers might be more invested in providing useful feedback, which can sometimes be more critical, leading to lower sentiment scores. Research has shown that reviewers with higher involvement or investment in the product or service are more likely to provide detailed and potentially critical feedback (see, e.g., Choi & Leon, 2020; Xu et al., 2022). Another could be associated with authentic reviewers feeling more motivated to provide detailed reviews that include both positive and negative aspects, leading to a lower overall sentiment score.

As a side analysis, and despite the non-normality of the dependent variable (sentiment score), additional linear regression models were performed. Results showed effects in different directions for the authentic coefficient when including or not the rating score variable. This suggests a potential confounding effect of rating scores. However, it's important to note that the validity of the linear regression results is compromised due to the violation of the normality assumption, underscoring the rigor of the research. While the observed effect direction is interesting, further analysis using other methods is needed for more precise conclusions.

Review length is another crucial factor in the study of online reviews, as it often reflects the effort and thoroughness put into the review. To ensure the validity of the current analysis,

[4] See Appendix 7.1 for details on the results.

atypical values were excluded. Therefore, all further analyses will consider only reviews with fewer than 2135 characters, which represents the 99th percentile, and larger than four characters, which represents the 1st percentile. This careful selection of data will allow for a more meaningful analysis by controlling for outliers and reviews with one character (a proper word is not made of one character). Furthermore, to maintain the normality assumption, a logarithmic transformation was implemented to the review length variable, further enhancing the robustness of the analysis.

ANOVA results indicate a statistically significant difference in review length between anonymous and authentic reviewers (F-value = 36140.31, p-value < 0.001). However, the very low Adj r-squared value, 0.003, suggests that the seeming status explains only a small fraction of the variance in review length. When running an ANOVA test without the log transformation, the results are consistent with an F-value of 25396.17 and a p-value <0.001. For a deeper understanding of the variables affecting review length, four different linear regression models were implemented. Table 11 presents results for Equation 3 following the same four variations implemented for Rating Scores and Sentiment Scores.

Equation 3: Review Length complete model

$$
\begin{aligned}
Log(Review\_Length_i) \\
&= \beta_0 + \beta_1 Authentic_i + \beta_2 Verified_i + \beta_3 Rating\_Score_i \\
&+ \sum_{j=1}^{28} \beta_{j+3} Category_i + \sum_{k=2}^{12} \beta_{k+31} Month_i + \varepsilon_i
\end{aligned}
$$

Notably, the log transformation applied to the dependent variable ensures that the residuals meet the normality assumption without issues, enhancing the reliability of the model. Further robustness checks with alternative models (Models A, B, and C) consistently show effects in the same direction with similar coefficients, reinforcing the stability and validity of the findings across different specifications. As a final robustness check, another linear regression was performed on the review length variable without any transformation. Despite the lack of validity, due to the non-normality when the transformation is not applied, the regression results presented as model E in Table 11, show consistent effect directions to what was already concluded from the other regression models.

Table 11: Review Length (characters number) Analysis

	DV: Review Length				
Model	(A)	(B)	(C)	(D)	(E)
Method	Linear Regression (Log Transformation)				Linear Regression
Authentic (binary)	-0.115***	-0.114***	-0.116***	-0.115***	-21.639***
	(0.673)	(0.001)	(0.001)	(0.001)	(0.154)
Verified	-0.969***	-0.972***	-0.875***	-0.878***	-240.609***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.330)
Rating Score	-0.138***	-0.138***	-0.136***	-0.136***	-19.672***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.601)
Fixed Effect (Month)		Included		Included	Included
Fixed Effect (Category)			Included	Included	Included
Observations	12,990,293	12,990,293	12,990,293	12,990,293	12,990,293
R-Squared	0.1003	0.1013	0.1281	0.1290	0.142

Note: ***p<0.01, **p<0.05, *p<0.1

As a final analysis, helpfulness votes have been studied as a cue for users when deciding their likelihood of purchase (see, e.g., Lei et al., 2023; Filieri et al., 2018; Xu et al., 2022). As previously explained, this condition is assigned entirely by external users. However, there is an interesting effect on analyzing how our independent variables could interact with anonymity and define the number of votes given to certain reviews. This analysis only considers reviews with at least one vote (vote > 0), and due to the skewness of the data, a logarithmic transformation is enhanced to meet the normality requirements.

The model considers seeming status, rating score, sentiment score, review length, and verified purchase as independent variables controlled by product and month-fixed effects[5]. Authenticity in reviewers, verified purchases, and rating scores have a negative effect on helpfulness votes, while review length and sentiment score have a positive impact. Results show that reviewers classified as authentic have, on average, 1.94% lower votes than anonymous reviewers. Interestingly, verified reviews receive significantly fewer votes compared to non-

[5] See Appendix 7.2 for details on the results.

verified reviews (coefficient = 0.188, p-value <0.001). Increases in rating scores could represent reductions of 2.9% in the number of helpfulness votes (p-value<0.001).

On the other hand, the longer the review, the more helpful votes it gets. Sentiment Score results suggest that more positive sentiment in reviews correlates with more votes. Overall, the regressions are statistically significant (p-value <0.000), explaining the 9.5% (R-squared) of the variation in helpfulness votes.

3.3 Discussion

The primary objective of this part of the study was to examine the influence of reviewer anonymity on review rating score, sentiment score, and length while using secondary data from Amazon. The different statistical analyses performed revealed various interesting findings. First, across all dependent variables (DVs), the study showed statistically significant differences between anonymous and authentic reviewers. The rating score is slightly higher for authentic reviewers. The sentiment score is lower for authentic reviewers, potentially due to a confounding effect of rating scores. Finally, authentic reviewers are associated with shorter reviews. It was consistently essential to include time and category fixed effects in the models, as these additions improved model performance in all cases.

Rating scores were analyzed as an ordinal variable, and ordinal logistic regression was used to control for non-normality. Although the proportional odds assumption was not perfectly met, various robustness checks, including multinomial logistic regression and additional linear regressions with transformations (as performed by Hoyer & Van Straaten, 2022), confirmed the general effects. Specifically, reviews from authentic-seeming names tend to have slightly higher ratings compared to anonymous ones. This suggests that perceived authenticity positively influences the ratings given by reviewers, aligning with H1 of this study and other authors' results (see, e.g., Deng et al., 2021; Ghasemaghaei et al., 2018).

In the same direction, sentiment scores have been proven to have a positive correlation with rating scores. This study's findings supported the same conclusion, finding a moderate correlation of 0.5107. Sentiment score, as a continuous variable, was analyzed through robust regressions to control for non-normality. Despite the positive correlation, an expected effect, results concluded that authentic reviewers are associated with lower sentiment scores, meaning more negative reviews. Supported by a multinomial logistic analysis, a possible explanation for this conclusion is that users feel more compromised when their "authentic" names are associated with the review, and this leads to a more partial review. Furthermore, this conclusion opens the possibility for further research on the effect of anonymity on different levels of sentiment score.

The third independent variable of interest was review length. Different linear regression analyses with a logarithmic transformation revealed that authentic reviewers write reviews that are approximately 11.52% shorter than those written by anonymous reviewers after controlling for rating scores, verified purchase, and fixed effects. This finding suggests that authentic reviewers tend to be more concise in their reviews compared to their anonymous counterparts, aligning with H3.

An additional analysis was performed regarding helpfulness votes, finding that authentic reviewers (compared to anonymous reviewers), verified purchases (compared to non-verified), and rating scores are, on average, associated with the lower number of helpfulness votes. On the other hand, review length (measured in the number of characters) and sentiment score showed a positive relation, increasing the number of helpfulness votes. The low R-Square suggests that other variables should be included in the analysis to better understand the variation of helpfulness votes.

These findings collectively highlight the nuanced ways in which authenticity influences reviewer behavior across different aspects of review content. The significant differences observed across all DVs emphasize the need for platforms and businesses to consider reviewer authenticity in their strategies. Future research should continue to explore these dynamics, particularly across different platforms and contexts, to further validate and extend these insights.

This part of this research has some limitations that need to be considered when generalizing the results. Firstly, since the data only comes from one platform, Amazon, it could lead to platform-specific results, requiring the inclusion of other platforms' data to generalize the results—secondary data. Furthermore, reviewers' data always has a strong component of self-selection bias since writing a review is entirely optional.

Thirdly, Amazon data was updated on the 6th of March 2023 to a more recent dataset containing data from 2018 until October 2023, but until the day of the presentation of this paper, the reviewer's name has not been included in the latest version, limiting the possibility of classifying the data. This limitation presents a possibility for further research when the reviewer's name is included as a possibility to analyze if there has been any significant change in the reviewer's behaviors over time.

One key factor in this analysis is the classification of reviewers' names between anonymous and authentic. Some package limitations have been extensively discussed regarding the use of the 'gender()' package. One of the main limitations of this package is its binary approach to gender, as it only considers females and males. However, due to the scope of this research, this is not an issue since gender is not a fundamental part of the analysis (see, e.g., Kozlowski et al., 2022; Lockhart et al., 2023). As presented by Mohammad (2020), another limitation is the exclusion of gender-fluid names and more recently used names.

4. Study 2: Anonymity when consuming content (Primary Data)

4.1 Survey Design

This research employed a 2 x 2 factorial within-subject experimental design to examine the proposed hypotheses and research question. In the study, two factors were systematically manipulated: the disclosure of the review sender's identity (authentic name vs. anonymous) and the sentiment expressed in the review (positive vs. negative). To maintain coherence with the first part of this study, the two most popular usernames identified previously as anonymous, as well as the two most popular authentic names, were used for the different reviews presented to the respondents.

Table 12: Survey Design - Review Situations

	Anonymous	**Authentic**
Positive	Review 1	Review 2
Negative	Review 3	Review 4

Research indicates that electronics represent the predominant category for seeking online reviews prior to making purchase decisions in contemporary consumer behavior (Freddie, 2018; Team, 2022). Consequently, to enhance engagement levels, earphones were selected as the focus of the survey. To avoid response bias and considering the within-subject design, four different products were chosen in order to fulfill the four reviews needed for the research presented in Table

12. Based on online technology opinion blogs, the best wireless earbuds for 2024 were chosen as the products to be included in the survey (Phelan, 2024).

Additionally, to avoid any research bias, the positive and negative reviews given for the product were taken from the Amazon website for the previously mentioned product. In order to prevent any misleading content or reviews that may exhibit both positive and negative sentiments, the most recent positive review, awarded five stars, was selected as the survey creation date (12th March 2024). Correspondingly, the negative review was deliberately chosen using the same criteria, focusing on the utmost extreme rating of 1 star available on the same previously mentioned date.

The questionnaire was created in Qualtrics and distributed through the researcher's personal social media accounts to maximize outreach and get more diverse responses. A power analysis related to the anonymity effect found across different contexts (Forman et al., 2008; Tsikerdekis, 2012; Hong & Cong, 2017; Pu et al., 2020; Deng et al., 2021; Hoyer & Van Straaten, 2022), determined that a minimum sample size of 80 participants was crucial for obtaining reliable results (AI-Therapy | Statistics for Psychologists | Sample Size Calculator, n.d.; Soper, n.d.)[6].

Before distributing the survey, a pre-test involving five participants was conducted to ensure that the usernames presented as anonymous and authentic were appropriately categorized by participants into the corresponding groups. Additionally, the pre-test was used to identify if any misinterpretation had taken place regarding the intention of the review (positive or negative) and to obtain an average duration time. As a result of the pre-test, some adjustments in the survey design were implemented. Review length was an issue for the average participant; this issue was resolved by choosing relatively shorter reviews. Additionally, the level of technical information included in the review was also highlighted by 3 participants to be too detailed, which "got them confused."; this was resolved by choosing more general reviews. Finally, a visual presentation was implemented to make Amazon's reviews look legitimate.

As commitment questions have been proven to provide better data quality results than attention checks on survey designs (Qualtrics, 2022), the first question included was a general commitment. Then, to determine eligibility for the survey, a screening question was asked at the

[6] See Appendix 7.3 for details.

beginning of the study: “Please indicate whether you would consider using online reviews when making technology purchases.” Participants who responded “NO” were then sent to the end of the survey, while participants who responded “YES” proceeded with the survey.

Due to the scope of the research, two main metrics were targeted: reliability of the review (content) and purchase intention. Kusumasondjaja et al. (2012) utilized a measurement scale for online perceived credibility borrowed from Flanagin and Metzger (2000). This scale assessed five key dimensions: accuracy, believability, unbiasedness, completeness, and trustworthiness. The previously presented scale was implemented in the survey, and participants were tasked with rating their agreement with these statements on a 7-point Likert scale, ranging from strongly disagree to strongly agree. Regarding purchase intention, many scales have been developed over time in marketing (MacKenzie et al., 1986; Holmes & Crocker, 1987; Spears & Singh, 2004; Bruner, 2019). However, for simplicity and due to the length of the survey, a three-item scale was used: likely/unlikely, probable/improbable, and possible/impossible, varying in a seven-point semantic scale (MacKenzie et al., 1986; Bruner, 2019).

Lastly, as recommended by Hughes, Camden, Yangchen, et al. (2016), since the primary objective of the survey was not to assess demographics as a critical factor, three demographic questions related to gender, age, and country of residency were placed at the end of the survey. This approach was employed to maintain participant interest in the study and mitigate survey fatigue, particularly in anticipation of more substantive inquiries[7].

4.2 Respondents

After the already explained pre-test, the final version of the survey was launched on March 27, 2024, and opened until April 13, 2024. A total of 180 responses were recorded, but only 135 successfully passed the qualification question. As the survey allowed respondents to drop at any point, out of those, 106 answered at least one of the scenarios. Finally, 88 respondents answered all questions related to the four scenarios. Figure 4 shows a summary funnel for respondents.

[7] Survey design is included in Appendix 7.4

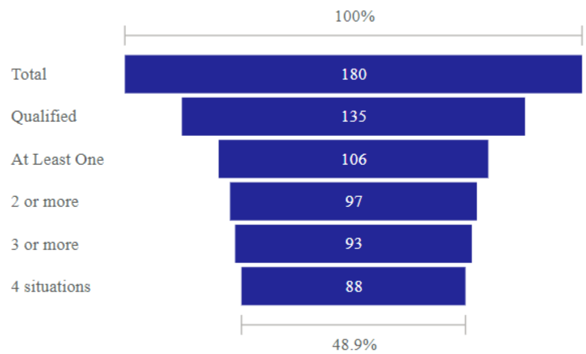

Figure 4: Respondents Funnel

Respondents answered the survey on their mobile phones in 75% of cases (79 respondents) compared to 25% who used a laptop. Regarding the duration of the study, the pre-test indicated an average duration of 5 minutes, while the actual survey had an average duration of 6 minutes, excluding atypical values.

Since demographic questions were asked at the end of the survey, only participants who completed all scenarios reached the demographic questions. Out of the 88 respondents, 51.1% identified as male, while 48.9% identified as female. Participants' ages ranged from 16 to 57 years old, with a median of 27 years old and an average of 29 years old. In terms of higher level of education completed, 38.6% (34) of respondents held a bachelor's degree, followed by 36.4% (32) with a master's degree. High school education was reported by 21.6% (19) of respondents, while other education levels, including PhD, were reported by only one respondent each.

Regarding the country of residency, Colombia had the highest representation with 34 respondents, followed by the Netherlands with 26. USA, UK, and Germany followed with 7, 5, and 4 respondents respectively. Other countries with two or fewer respondents were also included.

Due to the survey design, all participants faced all four scenarios. However, to control any order bias, the order of the scenarios was randomized. This led to an uneven number of responses in each scenario. Overall, 100 unique respondents answered questions facing an anonymous reviewer name compared to 101 facing an authentic reviewer name. Furthermore, 105 positive reviews compared to 98 with a negative sentiment. Reliability of the review and purchase intention were asked on a scale from 1 to 7, and on average, respondents rated 5.03 as reliability for anonymous reviewers compared to 4.80 when authentic. When comparing positive vs. negative sentiment reviews, the average for reliability was 5.17 and 4.64, respectively. Furthermore, purchase intention was also measured on a scale from 1 to 7, with a higher average for positive reviews, 5.22, compared to 2.96 for negative and a very similar average between anonymous and authentic reviews (4.01 for the first one and 4.24 for the latter one).

Table 13: Descriptive statistics primary data

Seeming Status	Sentiment	Rating	# Respondents	Avg Reliability	Avg Purchase Intention
Anonymous	Positive	5	97	5.42	5.27
	Negative	1	94	4.61	2.70
Authentic	Positive	5	100	4.93	5.17
	Negative	1	93	4.68	3.24

4.3 Results

This part of the research aims to analyze the effect of anonymity on user content consumption. Demographic variables collected during the survey will serve as control variables; therefore, the model will consider age, gender, the highest level of education completed, and current country of residency as controls. Furthermore, purchase intention is set as the main dependent variable, reviewer anonymity as the main independent variable, review reliability as a potential mediator, and review sentiment as a possible moderator.

Due to the nature of the data, multiple linear regression models were implemented to test different effects. Linear regression is a statistical method that tests the relationship between a dependent variable and one or more independent variables, fitting a linear equation with the data available. The main assumption for this model requires a linear relationship between the dependent variable and all independent variables. However, there are other important assumptions, such as independence, homoskedasticity, normality, exogeneity, and others (Poole & O'Farrell, 1970). The dataset compiled from the survey results successfully satisfies all of the assumptions, allowing us to use this method for further analysis.

First, the mediation effect of perceived reliability on the relationship between anonymity and purchase intention was tested. Baron and Kenny (1986) presented a methodology for testing a mediation hypothesis through linear regression analysis. The method intends to analyze mediation through three main linear regressions: firstly, testing how the independent variable predicts both the dependent variable and the potential mediator (on separate regression) and then how the independent variable and mediator predict the dependent variable. Following this methodology, Equation 4, Equation 5, and Equation 6 (presented below) were estimated.

Equation 4: Mediation - Linear Regression for IV (Authentic) effect on DV (Purchase Intention)

$$\begin{aligned} PurchaseIntention_i &= \beta_0 + \beta_1 Authentic_i + \beta_2 Positive_i + \beta_3 Age_i + \beta_4 Male_i + \sum_{j=1}^{4} \beta_{j+4} Education_i \\ &+ \sum_{k=1}^{14} \beta_{k+8} Country_i + \varepsilon_i \end{aligned}$$

Equation 5: Mediation - Linear Regression for IV (Authentic) effect on Mediator (Reliability)

$$\begin{aligned} Reliability_i &= \alpha_0 + \alpha_1 Authentic_i + \alpha_2 Positive_i + \alpha_3 Age_i + \alpha_4 Male_i + \sum_{j=1}^{4} \alpha_{j+4} Education_i \\ &+ \sum_{k=1}^{14} \alpha_{k+8} Country_i + \varepsilon_i \end{aligned}$$

Equation 6: Mediation - Learn Regression for IV (Authentic) and Mediator (Reliability) effect on DV (Purchase Intention)

$$\begin{aligned} PurchaseIntention_i &= \gamma_0 + \gamma_1 Authentic_i + \gamma_2 Reliability_i + \gamma_3 Positive_i + \gamma_4 Age_i + \gamma_5 Male_i \\ &+ \sum_{j=1}^{4} \gamma_{j+5} Education_i + \sum_{k=1}^{14} \gamma_{k+9} Country_i + \varepsilon_i \end{aligned}$$

The first regression, testing the direct effect of anonymity on purchase intention, has a positive and significant impact with a coefficient of 0.297 and a p-value of 0.021. Moving to the effect of reviewer anonymity on how reliable users found the review, there's a negative relation between both variables, with a coefficient of -0.219 and p-value of 0.046, indicating that reviews written by a reviewer considered authentic are perceived as less reliable than those written by anonymous reviewers.

Furthermore, the last regression, considering the effect of both the independent variable (IV) and mediator on the dependent variable (DV), has interesting implications indicating a competitive mediation effect. First, the coefficient for the IV authentic got stronger, going from 0.297 to 0.336, holding its statistical significance in both cases (p-value<0.001). Moreover, the mediator coefficient was estimated as 0.179 with a p-value of 0.014. The suggested competitive mediation effect shows that authenticity directly increases purchase intention but indirectly decreases it through reduced perceived reliability. In other words, while authentic reviews may initially boost purchase intention, their lower reliability perception partially offsets this effect. Figure 5 presents a summary of the different coefficients obtained when testing the model[8] and its significance based on ***p<0.01, **p<0.05 and *p<0.1.

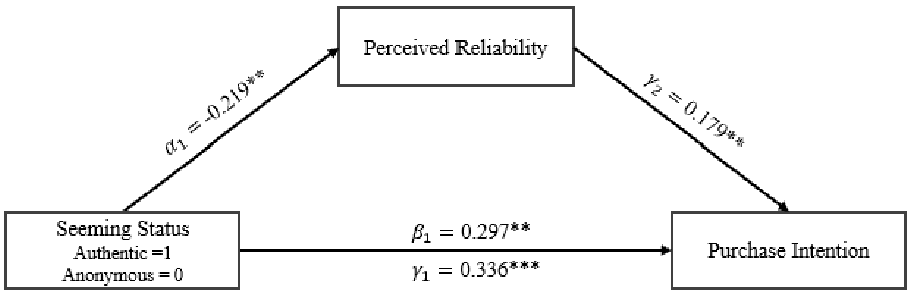

Figure 5: Mediation test of Reliability on Reviewer's Anonymity relationship with Purchase Intention

An interesting and unexpected finding from the previous models shows a negative relation between reviewers categorized as authentic and the perceived reliability, suggesting that reviews written by authentic reviewers are perceived as less reliable (α_1). Digging deeper into this relationship, it is interesting to check hypotheses five, which propose a moderation effect of review sentiment (positive or negative) in the relationship between reviewer anonymity and perceived reliability. To test this effect, a linear regression testing Equation 7 was performed. The interaction term was defined as Authentic * Positive.

Equation 7: Effect of Review Sentiment (Moderator) with Anonymity (IV) and Reliability (DV)

$$Reliability_i = \alpha_0 + \alpha_1 Authentic_i + \alpha_2 Positive_i + \alpha_3 Interaction_i + \alpha_4 Age_i + \alpha_5 Male_i + \sum_{j=1}^{4} \alpha_{j+5} Education_i + \sum_{k=1}^{14} \alpha_{k+9} Country_i + \varepsilon_i$$

When comparing the results of this new equation to those obtained in Equation 5, the R-squared value suggests that the model fits better when considering the moderation effect, with an R-squared of 0.2020 compared to 0.1875 in the previous model; both models are statistically

[8] Appendix 7.5 contains details for all coefficients estimated on each regression.

significant with p-values < 0.001. Table 14 presents a comparison between the regression model defined in Equation 5 without the moderation effect and Equation 7 with the interaction term[9].

Regarding the direction of the effects, when considering the moderation effect, the effect of being an authentic reviewer on reliability is positive but not statistically significant at the 0.05 level. This suggests that authenticity does not have a significant direct effect on perceived reliability in this model. Additionally, reviews with positive sentiment are perceived as more reliable. This effect is consistent across both regressions but becomes more pronounced when the interaction term is included, increasing the coefficient from 0.553 to 1.348. The negative and significant interaction term indicates that the effect of seeming status on reliability is moderated by sentiment. Specifically, the positive effect of being an authentic reviewer on reliability decreases when the sentiment is positive. In other words, the increase in perceived reliability due to positive sentiment is less pronounced for authentic reviewers compared to anonymous ones.

Table 14: OLS comparison with moderation effect on Reliability

DV: Reliability

Equation	(5)	(7)
Method	Linear Regression	
Authentic	-0.219**	0.575
	(0.109)	(0.356)
Positive (sentiment)	0.553***	1.348***
	(0.109)	(0.334)
Interaction		-0.53**
		(0.217)
Constant	4.008***	4.405***
	(0.467)	(0.475)
Control Variables	Included	Included
# Observations	352	352
R-Square	0.188	0.202
Prob > F	0.000	0.000

Note: ***p<0.01, **p<0.05, *p<0.1

To further explore the dynamics of user behavior, and due to the already significant effect found when including sentiment as a moderator, there is another possible moderation effect that is

[9] See Appendix 7.6 for details on the results.

analyzed: the moderation by sentiment on anonymity and purchase intention (Equation 8 presents the model). Since the possible moderation is being tested for the same moderator and IV, the definition of the interaction terms follows the same structure as previously defined for Equation 7.

Equation 8: Effect of Review Sentiment (Moderator) with Anonymity (IV) and Purchase Intention (DV)

$$Purchase\ Intention_i = \alpha_0 + \alpha_1 Reliability_i + \alpha_2 Authentic_i + \alpha_3 Positive_i + \alpha_4 Interaction_i + \alpha_5 Age_i + \alpha_6 Male_i + \sum_{j=1}^{4} \alpha_{j+6} Education_i + \sum_{k=1}^{14} \alpha_{k+10} Country_i + \varepsilon_i$$

Table 15: OLS comparison with moderation effect on Purchase Intention

DV: Purchase Intention

Equation	(6)	(8)
Method	Linear Regression	
Reliability	0.179**	0.158**
	(0.072)	(0.071)
Authentic	0.336***	1.249***
	(0.125)	(0.424)
Positive (sentiment)	2.216***	3.145***
	(0.139)	(0.388)
Interaction		-0.612**
		(0.251)
Constant	2.259***	2.801***
	(0.610)	(0.658)
Control Variables	Included	Included
# Observations	352	352
R-Square	0.535	0.543
Prob > F	0.000	0.000

Note: ***p<0.01, **p<0.05, *p<0.1

Once more, the model considering the moderation effect appears to be a better fit based on R-squared: the variables in Equation 8 explain 54.3% of the variation in purchase intention compared to 53.5% when the interaction is not considered. All variables have significant effects (p-value <0.001) and go in the same direction in both models. Reliability, which was proven to be a mediator for authenticity and purchase intention, shows a decrease in the effect magnitude from 0.179 to 0.158 when considering the interaction term. Overall, this suggests that higher perceived reliability of a review increases the likelihood of purchase, but in a smaller magnitude when interaction is considered[10].

When analyzing the variables involved in the interaction term, it is remarkable to notice that for both authentic and positive, the individual effects on purchase intention increased by almost 1 unit when including the moderation effect while maintaining high significance (p-value < 0.05). Furthermore, the interaction term suggests that the positive effect of being an authentic reviewer on purchase intention decreases when the sentiment is positive. In other words, the increase in purchase intention due to positive sentiment is less pronounced for reviews from authentic reviewers compared to anonymous ones.

The model defined in Equation 8 demonstrates the best fit among all the models tested, as evidenced by its higher R-squared value. This model effectively incorporates both mediation and moderation effects, providing a comprehensive understanding of how anonymity, reliability, and sentiment interact to influence purchase intention.

4.4 Discussion

The primary objective of this part of the study was to examine the influence of reviewer anonymity on purchase intention, either directly moderated by review sentiment or mediated by review reliability. The different statistical analyses performed revealed various interesting findings. First, there is a competitive mediation effect, where authenticity has both a direct positive effect and an indirect negative effect (through perceived reliability) on purchase intention. It means that, as stated in H4, authenticity increases purchase intention directly, but the negative impact of perceived lower reliability somewhat counteracts this effect. These results aligned with what was previously concluded by Forman (2008), Hong and Cong (2017), and Pavlou & Dimoka (2006), providing a possibility of including additional factors such as the seller's asymmetry of information for a future research, as other authors did.

The review sentiment has been found to have a significant effect on purchase intention. Therefore, to test for the implication of polarity sentiments, two moderation effect hypotheses were tested: one suggesting moderation of review sentiment on anonymity and reliability and another on anonymity and purchase. Both these relationships are key in this research and demonstrate

[10] See Appendix 7.7 for details on the results.

significant improvements in model fit when the interaction term is considered. Specifically, the interaction term indicated that sentiment alters the strength or direction of the relationship between seeming status and reliability. This result implies that while positive sentiment generally increases the perceived reliability of a review, this effect is less pronounced for reviews written by authentic reviewers, aligning with H5. Moreover, these findings are in line with what Filieri et al. (2018) concluded, but in particular, for authentic reviewers, Martin (2017) found a stronger effect when facing positive reviews compared to this paper's findings, suggesting a smaller effect for identity-revealing reviewers.

The second moderation effect tested is perhaps the most interesting one, as companies, most of the time, focus their strategies on increasing purchase intention. The analysis indicates that sentiment significantly moderates the relationship between anonymity (authenticity) and purchase intention (main DV). Furthermore, while both authenticity and positive sentiment individually boost purchase intention, the combination of authenticity and positive sentiment has a less harmonious effect than expected. Just as the mediation effect, this conclusion is also aligned with H6 and some authors, such as East et al. (2008) and Rosario et al. (2016).

This part of this research has some limitations in terms of data collection points, survey design, and methodology. As for the data collection points, there is a clear sample bias because the survey was mainly distributed through the researcher's social network. Overall, it cannot be assumed that the researcher's network is a significant representation of the overall review readers. This brings an opportunity for further research, expanding the sample size, and testing for consistent effects.

Moreover, the survey design contained two main limitations: first, as in any within-subject experiment, there is a likelihood of respondents getting what is called a "fatigue effect," meaning they start selecting random answers due to the length of the study. Second, the exposure to similar questions, in this case, four scenarios with the same questions, might result in either response biases or learning effects, revealing the purpose of the study and changing the naturality of the responses. Combining those limitations, considering a between-subject design, and controlling for any external effects using the same review on each scenario could allow for more general results.

Lastly, regarding the methodology, this study measured sentiment score as a binary variable based on polarity sentiments: either positive or negative. This approach could be enhanced by including a third neutral category or by creating a broader spectrum of sentiment possibilities. Such an improvement would allow for a more nuanced understanding of whether the effects observed are consistent not only for extreme emotions but also for less polarized sentiments.

5. Discussion

5.1 General Conclusions

This study contributes to the understating of the effect of anonymity within an online ecosystem. The goal of this study is to cover both content creation (written reviews) and content consumption (reading reviews) through different perspectives. To analyze content creation, the study presents a primary data section, with secondary data from Amazon online reviews for various product categories. This first study aims to cover the effect of anonymity on rating score, sentiment score, and review length. To explore content consumption, the study presents its own survey design with a within-subject experiment collecting primary data. This second study intends to cover the effect of anonymity on purchase intention while considering perceived reliability as a mediator and review sentiment as a moderator.

This paper highlights the importance of anonymity in analyzing online reviews for both content creation and content consumption. Rating score and sentiment score were found to be moderately correlated within a content production context. Answering research question (1), rating score itself was found to have a negative relation with anonymity, implying that anonymous reviewers give lower rating scores. However, sentiment score and review length (measured as a number of characters) were found to be higher when the reviewer was anonymous. The contradictory results between the rating score and sentiment score required further analysis, although initially, a possible explanation is associated with authentic reviewers feeling a higher sense of responsibility with their public content, leading to a more neutral or even negative review as it requires more honesty.

Within the context of content consumption (research question 2), it was found that authenticity increases purchase intention directly, but the negative impact of perceived lower reliability reduces this effect. The fact that the review was considered a positive or negative one moderated the relationship between perceived reliability and purchase intention. While a positive review increases the perceived reliability, the effect is weaker when an authentic reviewer writes the review. Furthermore, it was also concluded that while both a reviewer's authenticity and a positive review increase purchase intention, when combined, the effect is lower than expected. The dual impact highlights the complexity of user perceptions in online reviews. While authentic reviews may seem more trustworthy in terms of genuine user identity, they simultaneously reduce the perceived reliability of the reviews, thus influencing purchase intention in opposing ways.

Summarizing, it can be concluded that while anonymity encourages more expressive and extensive feedback, it may also result in harsher ratings. Conversely, within the context of content consumption, authenticity positively influences purchase intention, but the perceived lower reliability of authentic reviews mitigates this effect. Additionally, sentiment plays a key role in both contexts: content production and content consumption. Although measured on a different scale[11], sentiment score has a significant effect. Together, these findings highlight the dual-edged nature of anonymity in online reviews: while it raises more detailed and emotionally expressive content, it simultaneously presents challenges in balancing perceived reliability and user trust, ultimately impacting purchase decisions.

A managerial implication of these results lies in platforms recognizing the power of online reviews in influencing purchase intentions. Platforms often face challenges in understanding where their real issues lie. Based on the current results, it appears that allowing anonymity helps mitigate the positivity bias in rating scores. This strategy, however, can be a double-edged sword. On the one hand, sellers could receive a more realistic view of areas needing improvement and identify which products are underperforming. On the other hand, users reading these reviews might be influenced by the lower ratings, potentially affecting their purchase decisions. Balancing the benefits of honest feedback with the potential impact on user perceptions is crucial for platforms to consider.

5.2 Limitations and Future Research

As discussed in each section, some study-specific limitations apply to the current research. From a more global perspective, one of the biggest limitations is that the current study focuses explicitly on the effect of anonymity on online reviews, leaving other contexts out of the study. As

[11] Review sentiment is measured as a continuous variable between -1 (extremely negative) to 1 (extremely positive) for study 1. This is calculated based on a Natural Language Toolkit (NLTK), while it is measured as a binary variable (either positive or negative) for study 2.

a future research opportunity, researchers could analyze the impact of anonymity on social media behavior, support forums, or community sites.

Another limitation is the different timeframes considered in both studies. While Study 1, with secondary data collected from online reviews from 2008 until October 2018, study 2 collected data over three weeks from March 2024 until April 2024. This difference in timeline could lead to discrepancies in consumer behavior over time. Therefore, researchers could run both data collection points within the same timeframe to control for any time bias.

Lastly, both study 1 (secondary data) and study 2 (primary data) may have cultural or regional biases that limit the generalizability of the findings to other geographical contexts. In fact, study 2 heavily relies on Colombian respondents, followed by Dutch respondents, while study 1 features English-based text reviews, which are assumed to be more representative of the US or UK. Another possibility for future research could be running both studies within the same geographical context to either get country/region-specific results or global results with a representative sample of different countries for both studies.

6. References

About public names - Amazon customer service. (n.d.). https://www.amazon.in/gp/help/customer/display.html?nodeId=202111170

AI-Therapy | Statistics for Psychologists | Sample size calculator. (n.d.). https://www.ai-therapy.com/psychology-statistics/sample-size-calculator

Amazon Verified Purchase Reviews. (n.d.). Amazon. https://www.amazon.com/gp/help/customer/display.html?nodeId=G75XTB7MBMBTXP6W

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

Ashtiani, M. N., & Raahemi, B. (2023). News-based intelligent prediction of financial markets using text mining and machine learning: A systematic literature review. *Expert Systems With Applications, 217*, 119509. https://doi.org/10.1016/j.eswa.2023.119509

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

Bruner, G. C. (2019). Marketing Scales Handbook: Multi-Item Measures for Consumer Insight Research (Volume 10). *ResearchGate*, *10*.

Bruner, G. C. (2021). Marketing Scales Handbook: Multi-Item Measures for Consumer Insight Research (Volume 11). *ResearchGate, 11*.

Can I review anonymously at Amazon? (2016, December 4). Kindle Forum. https://www.kboards.com/threads/can-i-review-anonymously-at-amazon.244658/

Chen, P.-Y., Hong, Y., & Liu, Y. (2018). The Value of Multidimensional Rating Systems: Evidence from a Natural Experiment and Randomized Experiments. *Management Schience*, *64–64*(10), 4629–4647. http://kevinhong.miami/paper-pdf/MS_MD_Systems_2018.pdf

Chen, W., Gu, B., Ye, Q., & Zhu, K. (2019). Measuring and managing the externality of managerial responses to online customer reviews. *Information Systems Research*, *30*(1), 81–96. https://doi.org/10.1287/isre.2018.0781

Choi, H., & Leon, S. (2020). An empirical investigation of online review helpfulness: A big data perspective. *Decision Support Systems*, *139*, 113403. https://doi.org/10.1016/j.dss.2020.113403

Cohen's kappa. (2024). DataTab. https://datatab.net/tutorial/cohens-kappa

Deng, L., Sun, W., Xu, D., & Ye, Q. (2021). Impact of anonymity on consumers' online reviews. *Psychology & Marketing*, *38*(12), 2259–2270. https://doi.org/10.1002/mar.21565

Diamantini, C., Mircoli, A., Potena, D., & Storti, E. (2019). Social information discovery enhanced by sentiment analysis techniques. *Future Generation Computer Systems*, *95*. https://www-sciencedirect-com.eur.idm.oclc.org/science/article/pii/S0167739X17308634

Dixon, S. J. (2024, February 29). *Topic: Online reviews*. Statista. https://www.statista.com/topics/4381/online-reviews/

East, R., Hammond, K., & Lomax, W. (2008). Measuring the impact of positive and negative word of mouth on brand purchase probability. *International Journal of Research in Marketing*, *25*(3), 215–224. https://doi.org/10.1016/j.ijresmar.2008.04.001

East, R., Hammond, K., & Wright, M. (2007). The relative incidence of positive and negative word of mouth: A multi-category study. *International Journal of Research in Marketing*, *24*(2), 175–184. https://doi.org/10.1016/j.ijresmar.2006.12.004

Encyclopedia of Social Measurement. (2005). ScienceDirect. https://www.sciencedirect.com/referencework/9780123693983/encyclopedia-of-social-measurement

Fan, W., Yong, L., Li, H., Tuunainen, V. K., & Lin, Y. (2021). Quantifying the effects of online review content structures on hotel review helpfulness. *Internet Research*, *32*(7), 202–227. https://doi.org/10.1108/intr-11-2019-0452

Filieri, R., Galati, F., & Raguseo, E. (2021). The impact of service attributes and category on eWOM helpfulness: An investigation of extremely negative and positive ratings using latent semantic analytics and regression analysis. *Computers in Human Behavior, 114*, 106527. https://doi.org/10.1016/j.chb.2020.106527

Filieri, R., Raguseo, E., & Vitari, C. (2018). What moderates the influence of extremely negative ratings? The role of review and reviewer characteristics. *International Journal of Hospitality Management*. https://www.sciencedirect.com/science/article/abs/pii/S0278431918302500

Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The Role of Reviewer Identity Disclosure in Electronic Markets on JSTOR. *www.jstor.org*. https://www.jstor.org/stable/23015450

Freddie. (2018, May 17). *Revealed - Top categories affected by online customer reviews - Product Review Monitoring*. Product Review Monitoring. https://reviewmonitoring.com/revealed-top-categories-affected-by-online-customer-reviews/#:~:text=Shoppers%20looking%20to%20purchase%20electronic,before%20making%20a%20purchase%20decision.

Ghasemaghaei, M., Eslami, S. P., Deal, K., & Hassanein, K. (2018). Reviews' length and sentiment as correlates of online reviews' ratings. *Internet Research*, *28*(3), 544–563. https://doi.org/10.1108/intr-12-2016-0394

Gutt, D., & Neumann, J. (2019). *THE VIRTUES OF ANONYMITY – AN EMPIRICAL INVESTIGATION OF THE RELATIONSHIP BETWEEN B2B ONLINE RATINGS AND REVIEWER SELF-DISCLOSURE*. AIS Electronic Library (AISeL). https://aisel.aisnet.org/ecis2019_rip/29/

Ho, J. Y., & Dempsey, M. (2010). Viral marketing: Motivations to forward online content. *Journal of Business Research*, *63*(9–10), 1000–1006. https://doi.org/10.1016/j.jbusres.2008.08.010

Holmes, J., & Crocker, K. E. (1987). Predispositions and the comparative effectiveness of rational, emotional and discrepant appeals for both high involvement and low involvement products. *Journal of the Academy of Marketing Science*, *15*(1), 27–35. https://doi.org/10.1007/bf02721951

Homburg, C., Ehm, L., & Artz, M. (2015). Measuring and Managing Consumer Sentiment in an Online Community Environment. *Journal of Marketing Research*, *52*(5). https://www.jstor.org/stable/43832390

Hong, C., & Cong, L. (2017). The Effect of "Anonymous Reviewer": A study of anonymity, affect intensity, and message valence in the cyberspace. *Journal of Language and Social Psychology*, *36*(5), 504–524. https://doi.org/10.1177/0261927x17697358

Hoyer, B., & Van Straaten, D. (2022). Anonymity and self-expression in online rating systems—An experimental analysis. *Journal of Behavioral and Experimental Economics*, *98*, 101869. https://doi.org/10.1016/j.socec.2022.101869

Huang, N., Hong, Y., & Burtch, G. (2016, November 18). *Social Network Integration and User Content Generation: Evidence from Natural Experiments*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2872275

Hughes, J. L., Camden, A. A., & Yangchen, T. (2016). Rethinking and Updating Demographic Questions: Guidance to improve descriptions of research samples. *Psi Chi Journal of Psychological Research*, *21*(3), 138–151. https://doi.org/10.24839/2164-8204.jn21.3.138

Hughes, J. L., Camden, A. A., Yangchen, T., Lorona, R. T., Miller-Perrin, C., Raiffe, D. E., Murphy, T. P., Keres, A., Chartier, C. R., Abel, E. A., Gadomski, M., Brodhead, M. T., Morris, W. L., Slonim, G., Osburn, B. K., Taggart, T. C., Hammett, J. F., Ulloa, E. C., Wright, R. R., . . . Yangchen, T. (2016). Psi Chi Journal of Psychological Research. *Psi Chi Journal of Psychological Research*, *21*(3), 138–200. https://cdn.ymaws.com/www.psichi.org/resource/resmgr/journal_2016/21_3Fall16JN-Hughes.pdf

Jiang, Y., Ho, Y., Yan, X., & Tan, Y. (2022). What's in a "Username"? The effect of perceived anonymity on herding in crowdfunding. *Information Systems Research*, *33*(1), 1–17. https://doi.org/10.1287/isre.2021.1049

Kozlowski, D., Murray, D., Bell, A. T., Hulsey, W., Larivière, V., Monroe-White, T., & Sugimoto, C. R. (2022). Avoiding bias when inferring race using name-based approaches. *PloS One*, *17*(3), e0264270. https://doi.org/10.1371/journal.pone.0264270

Kusumasondjaja, S., Shanka, T., & Marchegiani, C. (2012). Credibility of online reviews and initial trust. *Journal of Vacation Marketing*, *18*(3), 185–195. https://doi.org/10.1177/1356766712449365

Lei, Z., Yin, D., & Zhang, H. (2023). *Positive or negative reviews? Consumers' selective exposure in seeking and evaluating online reviews*. AIS Electronic Library (AISeL). https://aisel.aisnet.org/jais/vol24/iss4/1/

Lenhard, W. (2022). *Computation of different effect sizes like d, f, r and transformation of different effect sizes: Psychometrica*. https://www.psychometrica.de/effect_size.html

Lockhart, J. W., King, M. M., & Munsch, C. (2023). Name-based demographic inference and the unequal distribution of misrecognition. *Nature Human Behaviour*. https://www-nature-com.eur.idm.oclc.org/articles/s41562-023-01587-9.pdf

MacKenzie, S. B., Lutz, R. J., & Belch, G. E. (1986). The Role of Attitude toward the Ad as a Mediator of Advertising Effectiveness: A Test of Competing Explanations on JSTOR. *www.jstor.org*, *23*. https://www.jstor.org/stable/3151660?seq=1

Martin, W. C. (2017, November 20). *Positive versus negative word-of-mouth: Effects on receivers*. https://www.abacademies.org/articles/positive-versus-negative-wordofmouth-effects-on-receivers-6732.html#:~:text=The%20impact%20of%20receivers'%20using,their%20attitude%20toward%20the%20product.

Matsakis, L. (2019, May 25). What do Amazon's star ratings really mean? *WIRED*. https://www.wired.com/story/amazon-stars-ratings-calculated/

McAuley, J. (2019). *Recommender Systems datasets*. https://cseweb.ucsd.edu/~jmcauley/datasets.html#amazon_reviews

McHugh, M. L. (2012, October 15). *Interrater Reliability: The Kappa Statistic*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/#:~:text=Cohen%20suggested%20the%20Kappa%20result,1.00%20as%20almost%20perfect%20agreement.

Mohammad, S. M. (2020). Gender Gap in Natural Language Processing Research: Disparities in Authorship and Citations. *ACL*. https://doi.org/10.18653/v1/2020.acl-main.702

Mullen, L. (2021, October 12). *Predicting gender using historical data*. https://cran.r-project.org/web/packages/gender/vignettes/predicting-gender.html

Ni, J., Li, J., & McAuley, J. (2019). Justifying Recommendations using Distantly-Labeled Reviews and Fine-Grained Aspects. *EMNLP*. https://doi.org/10.18653/v1/d19-1018

NLTK :: Release Notes. (2023). https://www.nltk.org/news.html

Pan, X., Hou, Y., & Wang, Q. (2023). Are we braver in cyberspace? Social media anonymity enhances moral courage. *Computers in Human Behavior*, *148*, 107880. https://doi.org/10.1016/j.chb.2023.107880

Pavlou, P. A., & Dimoka, A. (2006). The nature and role of feedback text comments in online marketplaces: implications for trust building, price premiums, and seller differentiation. *Information Systems Research*, *17*(4), 392–414. https://doi.org/10.1287/isre.1060.0106

Peña-García, N., Gil-Saura, I., Rodríguez-Orejuela, A., & Siqueira-Junior, J. R. (2020). Purchase intention and purchase behavior online: A cross-cultural approach. *Heliyon*, *6*(6), e04284. https://doi.org/10.1016/j.heliyon.2020.e04284

Phelan, D. (2024, January 22). Best wireless earbuds 2024, tried and tested. *The Independent*. https://www.independent.co.uk/extras/indybest/gadgets-tech/headphones-earphones/best-wireless-earbuds-b422237.html

Pinsonneault, A., & Heppel, N. (1997). Anonymity in Group Support Systems Research: A new Conceptualization, Measure, and Contingency Framework on JSTOR. *www.jstor.org*. https://www.jstor.org/stable/40398279

Poole, M. A., & O'Farrell, P. N. (1970). The assumptions of the linear regression model on JSTOR. *www.jstor.org*. https://www.jstor.org/stable/621706?casa_token=lRrzFThKaXEAAAAA%3AU9TadTjk3u5OJ8qISTOjda5uWFWKEhOHssIepIFvArZxkaTc-ngl0GnyIj4NIEGwzEs8dBWGab86OfDx3qBPcxlIYO18TPnnbh1J62bCuwnt09ALfV4v&seq=7

Pu, J., Yuan, C., Qiu, L., & Cheng, H. K. (2020). Does identity disclosure help or hurt user content generation? Social presence, inhibition, and displacement effects. *Information Systems Research*, *31*(2), 297–322. https://doi.org/10.1287/isre.2019.0885

Qualtrics. (2022, August 29). *Improve data quality by using a commitment request instead of attention checks*. Qualtrics. https://www.qualtrics.com/blog/attention-checks-and-data-quality/

Rainie, L. (2005, January 1). *Internet: The mainstreaming of online life*. https://search.issuelab.org/resource/internet-the-mainstreaming-of-online-life.html

Rosario, A. B., Sotgiu, F., De Valck, K., & Bijmolt, T. H. A. (2016). The Effect of Electronic Word of Mouth on Sales: A Meta-Analytic Review of platform, product, and metric factors. *Journal of Marketing Research*, *53*(3), 297–318. https://doi.org/10.1509/jmr.14.0380

Schoenmueller, V., Netzer, O., & Stahl, F. (2020). The polarity of online reviews: Prevalence, Drivers and Implications. *Journal of Marketing Research*, *57*(5). https://www.jstor.org/stable/27181455

Soper, D. (n.d.). *Free A-priori sample size calculator for Student T-Tests - Free Statistics calculators*. https://www.danielsoper.com/statcalc/calculator.aspx?id=47

Spears, N., & Singh, S. (2004). Measuring Attitude toward the Brand and Purchase Intentions. *Journal of Current Issues and Research in Advertising*, *26*(2), 53–66. https://doi.org/10.1080/10641734.2004.10505164

Team, P. (2022, March 9). *Beyond the basics: key information shoppers look for in product reviews*. PowerReviews. https://www.powerreviews.com/insights/information-shoppers-want-product-reviews/

Teubner, T., & Glaser, F. (2018). UP OR OUT— THE DYNAMICS OF STAR RATING SCORES ON AIRBNB. *Association for Information Systems AIS Electronic Library (AISeL)*.

Tsikerdekis, M. (2012). The choice of complete anonymity versus pseudonymity for aggression online. *ResearchGate*. https://www.researchgate.net/publication/220023480_The_choice_of_complete_anonymity_versus_pseudonymity_for_aggression_online

Understanding Statistical Power and Significance testing — an interactive visualization. (n.d.). https://rpsychologist.com/d3/nhst/

Vana, P., & Lambrecht, A. (2021). The effect of individual online reviews on purchase likelihood. *Marketing Science*, *40*(4), 708–730. https://doi.org/10.1287/mksc.2020.1278

Wang, Y., & Chaudhry, A. (2018). When and How Managers' Responses to Online Reviews Affect Subsequent Reviews. *Journal of Marketing Research*, *55*(2). https://www.jstor.org/stable/44878549

What is named entity recognition? | IBM. (2022). https://www.ibm.com/topics/named-entity-recognition

Xu, D., Ye, Q., Hong, H., & Sun, F. (2022). Emotions for attention in online consumer reviews: the moderated mediating role of review helpfulness. *Industrial Management + Data Systems/Industrial Management & Data Systems*, *122*(3), 729–751. https://doi.org/10.1108/imds-07-2021-0473

Zhang, K., Kizilcec, R. F., Department of Communication, & Stanford University. (2014). Anonymity in social media: effects of content controversiality and social endorsement on sharing behavior. In *Association for the Advancement of Artificial Intelligence*. https://rene.kizilcec.com/wp-content/uploads/2014/03/zhang_kizilcec_anonymity_icwsm2014.pdf

Zhang, Z., Liang, S., Li, H., & Zhang, Z. (2019). Booking now or later: Do online peer reviews matter? *International Journal of Hospitality Management*, *77*, 147–158. https://doi.org/10.1016/j.ijhm.2018.06.024

Zhang, Z., Zhang, Z., & Zhang, Z. (2023). Why do online reviewers seek anonymity? Empirical evidence on the effects of prior anonymous reviews and focal reviews. *Aslib Proceedings*. https://doi.org/10.1108/ajim-08-2022-0369

Zimmermann, S., Herrmann, P., Kundisch, D., & Nault, B. R. (2018). Decomposing the variance of consumer ratings and the impact on price and demand. *Information Systems Research*, *29*(4), 984–1002. https://doi.org/10.1287/isre.2017.0764

7. Appendix

7.1 Multinomial Results – Study 1 Sentiment Score

Multinomial logistic regression was performed considering sentiment score >0 or positive score as the comparison group.

A. Sentiment Score <0 – Negative Score

sentiment_categ	Coefficient	Std. err	P>|z|	[95% conf. interval]	
0 = negative					
seeming_status_binary					
Authentic	-0.009	0.002	0.000	-0.013	-0.006
verified_binary					
TRUE	0.107	0.003	0.000	0.101	0.112
overall	-0.918	0.001	0.000	-0.919	-0.916
category_binary					
All_Beauty	0.324	0.008	0.000	0.309	0.339
Appliances	0.711	0.007	0.000	0.698	0.725
Arts_Crafts_and_Sewing	0.311	0.007	0.000	0.297	0.325
Automotive	0.582	0.007	0.000	0.569	0.596
Books	0.469	0.007	0.000	0.455	0.483
CDs_and_Vinyl	0.021	0.008	0.008	0.005	0.037
Cell_Phones_and_Accessories	0.475	0.007	0.000	0.461	0.488
Clothing_Shoes_and_Jewelry	-0.025	0.007	0.001	-0.040	-0.011
Digital_Music	0.077	0.008	0.000	0.061	0.094
Electronics	0.521	0.007	0.000	0.507	0.534
Gift_Cards	-0.158	0.014	0.000	-0.186	-0.131
Grocery_and_Gourmet_Food	0.229	0.007	0.000	0.214	0.243
Home_and_Kitchen	0.260	0.007	0.000	0.246	0.274
Industrial_and_Scientific	0.647	0.007	0.000	0.633	0.660
Kindle_Store	0.345	0.007	0.000	0.331	0.360
Luxury_Beauty	0.204	0.007	0.000	0.190	0.219
Magazine_Subscriptions	-0.088	0.014	0.000	-0.115	-0.061
Movies_and_TV	0.512	0.007	0.000	0.499	0.526
Musical_Instruments	0.315	0.007	0.000	0.301	0.329
Office_Products	0.465	0.007	0.000	0.452	0.479
Patio_Lawn_and_Garden	0.579	0.007	0.000	0.566	0.593
Pet_Supplies	0.466	0.007	0.000	0.453	0.480
Prime_Pantry	0.458	0.007	0.000	0.444	0.473
Software	0.400	0.007	0.000	0.387	0.414
Sports_and_Outdoors	0.338	0.007	0.000	0.324	0.352
Tools_and_Home_Improvement	0.506	0.007	0.000	0.493	0.520
Toys_and_Games	0.181	0.007	0.000	0.167	0.195
Video_Games	0.392	0.007	0.000	0.379	0.406

month						
	2	0.032	0.005	0.000	0.023	0.040
	3	0.032	0.004	0.000	0.023	0.040
	4	0.048	0.005	0.000	0.039	0.057
	5	0.041	0.005	0.000	0.032	0.050
	6	0.035	0.005	0.000	0.026	0.044
	7	0.037	0.005	0.000	0.028	0.046
	8	0.040	0.005	0.000	0.031	0.049
	9	0.041	0.005	0.000	0.032	0.050
	10	0.030	0.005	0.000	0.021	0.039
	11	0.040	0.005	0.000	0.031	0.049
	12	0.021	0.004	0.000	0.013	0.030
_cons		1.046	0.007	0.000	1.033	1.059

B. Sentiment Score =0 – Neutral Score

sentiment_categ	Coefficient	Std. err	P>|z|	[95% conf. interval]	
1 = neutral					
seeming_status_binary					
Authentic	0.085	0.002	0.000	0.081	0.090
verified_binary					
TRUE	0.868	0.004	0.000	0.860	0.876
overall	-0.488	0.001	0.000	-0.490	-0.487
category_binary					
All_Beauty	0.069	0.008	0.000	0.053	0.085
Appliances	0.489	0.007	0.000	0.476	0.503
Arts_Crafts_and_Sewing	0.253	0.007	0.000	0.239	0.267
Automotive	0.297	0.007	0.000	0.283	0.310
Books	-0.471	0.009	0.000	-0.488	-0.454
CDs_and_Vinyl	-0.422	0.009	0.000	-0.440	-0.405
Cell_Phones_and_Accessories	-0.039	0.007	0.000	-0.053	-0.024
Clothing_Shoes_and_Jewelry	-0.097	0.007	0.000	-0.112	-0.083
Digital_Music	0.056	0.008	0.000	0.041	0.071
Electronics	0.044	0.007	0.000	0.029	0.058
Gift_Cards	0.116	0.011	0.000	0.094	0.138
Grocery_and_Gourmet_Food	-0.046	0.008	0.000	-0.061	-0.031
Home_and_Kitchen	-0.095	0.008	0.000	-0.110	-0.081
Industrial_and_Scientific	0.559	0.007	0.000	0.545	0.572
Kindle_Store	-0.570	0.009	0.000	-0.588	-0.553
Luxury_Beauty	-0.161	0.008	0.000	-0.176	-0.145
Magazine_Subscriptions	-0.032	0.014	0.027	-0.060	-0.004
Movies_and_TV	-0.228	0.008	0.000	-0.244	-0.212
Musical_Instruments	-0.107	0.008	0.000	-0.122	-0.092

Office_Products	0.259	0.007	0.000	0.245	0.272
Patio_Lawn_and_Garden	0.272	0.007	0.000	0.258	0.286
Pet_Supplies	-0.102	0.008	0.000	-0.117	-0.087
Prime_Pantry	0.469	0.007	0.000	0.454	0.483
Software	0.017	0.008	0.029	0.002	0.032
Sports_and_Outdoors	0.001	0.007	0.848	-0.013	0.016
Tools_and_Home_Improvement	0.179	0.007	0.000	0.165	0.193
Toys_and_Games	-0.199	0.008	0.000	-0.214	-0.183
Video_Games	-0.191	0.008	0.000	-0.207	-0.176
month					
2	0.039	0.005	0.000	0.029	0.049
3	0.065	0.005	0.000	0.055	0.074
4	0.048	0.005	0.000	0.038	0.058
5	0.031	0.005	0.000	0.021	0.041
6	0.043	0.005	0.000	0.033	0.053
7	0.117	0.005	0.000	0.107	0.126
8	0.138	0.005	0.000	0.128	0.147
9	0.132	0.005	0.000	0.122	0.142
10	0.117	0.005	0.000	0.107	0.127
11	0.096	0.005	0.000	0.086	0.106
12	0.038	0.005	0.000	0.028	0.048
_cons	-1.212	0.008	0.000	-1.227	-1.197

7.2 Linear Regression – Study 1 Helpfulness Votes

	DV: Helpfulness Votes			
Model	(A)	(B)	(C)	(D)
Method	Linear Regression (Log Transformation)			
Authentic (binary)	-0.022***	-0.022***	-0.020***	-0.019***
	(0.001)	(0.001)	(0.001)	(0.001)
Verified	-0.173***	-0.172***	-0.189***	-0.188***
	(0.002)	(0.002)	(0.002)	(0.002)
Rating Score	-0.361***	-0.036***	-0.029***	-0.029***
	(0.000)	(0.000)	(0.000)	(0.000)
Review Length	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Sentiment Score	0.037***	0.037***	0.042***	0.042***
	(0.001)	(0.001)	(0.001)	(0.001)
Fixed Effect (Month)		Included		Included
Fixed Effect (Category)			Included	Included
Observations	1,827,782	1,827,782	1,827,782	1,827,782
Adj R-Squared	0.0843	0.0846	0.0945	0.0949

Note: ***p<0.01, **p<0.05, *p<0.1

7.3 Power Analysis – Survey

To estimate the ideal sample size for the survey design, a power analysis was performed. Considering different effect sizes from other studies that covered the effect of anonymity in content consumption across different industries and employing different methodologies. Effect sizes were calculated depending on the statistical method implemented within the paper. Some papers are required to calculate Cohen D, Cohen F, effect size coefficient w, etc. Additionally, since one paper could have different models that measure the effect of anonymity, a paper could have more than one effect size. Summarizing, the following effect sizes were found:

A. Deng et al. (2021)
    - Effect Size: 0.125
    - Effect Size: 0.193
B. Forman et al. (2008)
    - Effect Size: 0.170
C. Hong and Cong (2017)
    - Effect Size: 0.520
D. Hoyer and Van Straaten (2022)
    - Effect Size: 0.115
    - Effect Size: 0.202
E. Pu et al. (2020)
    - Effect Size: 1.137
    - Effect Size: 0.154
    - Effect Size: 0.147
    - Effect Size: 0.233
F. Tsikerdekis (2012)
    - Effect Size: 0.211
    - Effect Size: 0.137

The average effect size was calculated as 0.279. Using the sample size calculator for a within-subject design with an alpha of 0.05 a power of 0.8 and the average effect size found, the sample size is set as 80 (AI-Therapy | Statistics for Psychologists | Sample Size Calculator, n.d.).

7.4 Survey Design

Introduction

Thank you for participating in this survey; your valuable input is crucial for the success of this research project. Your participation in this survey is entirely voluntary. You have the right to withdraw at any point during the survey without providing a reason. Your responses will be treated with strict confidentiality. No personally identifiable information will be disclosed in any reports or publications resulting from this research. Data will be aggregated and anonymized to ensure your privacy. All data collected will be securely stored and accessible only to the researcher and authorized personnel. The survey platform used employs industry-standard security measures to protect the information you provide.

By continuing with this survey, you indicate your informed consent to participate. If you have any questions or concerns about the study, please get in touch with Carolina Rueda at 687807jr@eur.nl.

Question 1 - Commitment Question

We care about the quality of our survey data. To get the most accurate measure of your opinions, it is important that you provide thoughtful answers to each question in this survey. Your response is anonymous and will be treated with the highest level of confidentiality.

Do you commit to providing thoughtful answers to the questions in this survey?

- I can't promise either way.
- Yes, I will.
- No, I will not.

Question 2 - Screening Question

The upcoming questions involve a hypothetical scenario related to online reviews for technology and your purchasing intentions. Consequently, if you have no intention of utilizing online reviews in your decision-making process for technology purchases, the survey concludes at this point for you.

Please indicate whether you would consider using online reviews when making technology purchases:

- Yes, I would consider using online reviews when making technology purchases.
- No, I would not consider using online reviews when making technology purchases.

Disclaimer for hypothetical situations

Imagine you are contemplating the purchase of a new set of earphones. For the purpose of this survey, kindly respond to the following questions based solely on the reviews provided. Assume that you are genuinely interested in the product and refrain from incorporating your personal opinions or pre-existing knowledge about the product into your responses.

Positive – Anonymous Review (Review 1)


Sony WF-1000XM5
Amazon Customer
Reviewed in the Netherlands on March 7, 2024
Color: Black | Verified Purchase
I picked up these headphones because they are well rated everywhere, and I have to say I wasn't disappointed. The sound is good, but the clear distinguishing features are the additional functions and voice recognition. Everything works great. Only the battery life could be a bit longer, but with an average of 5 to 6 hours, it is also sufficient.

Question 3 (Reliability on the review)

Based on the previous review, please indicate how you feel about the following statements:

(Scale strongly disagree to strongly agree with 7 likert-scale)

The review is accurate
The review is believable
The review is unbiased
The review is complete
The review is trustworthy

Question 4 (Purchase Intention)

Based only on the previous review and setting aside any preconceived notions, please indicate your willingness of purchasing the product:

(7 likert-scale)

Unlikely (1) / likely (7)
Improbable (1) / Probable (7)
Impossible (1) / Possible (7)

Positive – Authentic Review (Review 2)

Bowers & Wilkins PI7

Mike

★★★★★

Reviewed in the Netherlands on March 7, 2024

Color: Black | **Verified Purchase**

The noise cancellation is spectacular, and the sound quality is beyond what apple AirPods offer. They don't fall out near as much as any other earphone I've had. I've had beats, AirPods, multiple cheaper wireless knockoffs and now these. You won't regret this decision! I love love listening to music with these in. It's like I'm zoned in better than ever before.

Question 5 (Reliability on the review)

Sames as question 3

Question 6 (Purchase Intention)

Sames as question 4

Negative – Anonymous Review (Review 3)

Bose - QuietComfort Earbuds II

Kindle Customer

★☆☆☆☆

Reviewed in the Netherlands on March 7, 2024

Color: Black | **Verified Purchase**

Had the issues with this product been limited to their basically non-existent noise cancelling capabilities, that render them useless on anything even resembling a busy street, I'd gladly call them overpriced. However, that is only where the problems start. They frequently simply refuse to connect to any device, they give off very weird echo artifacts at times, and the list just goes on.

Question 7 (Reliability on the review)

Sames as question 3

Question 8 (Purchase Intention)

Sames as question 4

Negative – Authentic Review (Review 4)

Beats Powerbeats Pro

Chris

Reviewed in the Netherlands on March 7, 2024

Color: Black | **Verified Purchase**

I have sent my 9 month old Powerbeats for a repair because the right one would not hold a charge. Amazon repair center "Ingram" declined to repair due to "liquid damage" - for heaven's sake they are marketed as Sweat-resistant. They have not been submerged and regularly dried after putting back in the case. Nice kit - but as it turns out not fit for the purpose.

Question 9 (Reliability on the review)

Sames as question 3

Question 10 (Purchase Intention)

Sames as question 4

Demographic Questions

Please choose the option that suits you the most:

D Question 1

What is your gender?

Female
Male
Non-binary / third gender
Prefer not to say

D Question 2

What is your age? (In years)

____

D Question 3

What is the highest level of education you have completed?

High School

Bachelor (University Degree)
Master
PhD
Other

D Question 4

Netherlands
Germany
Colombia
Other:________________

Closure

Thank you for taking the time to participate in this research. The survey ends here.

7.5 Mediation Test – Study 2

Barron & Kenny method to test mediation

A. Effect of IV on DV

purchase_intention	Coefficient	std. err.	P>|t|
seeming_status_binary			
Authentic	0.297	0.127	0.021
sentiment_binary			
Positive	2.315	0.127	0.000
age	-0.008	0.010	0.417
gender_2			
Male	0.219	0.140	0.119
education_2			
High-School	-0.314	0.183	0.088
Master	0.356	0.186	0.057
Other	-0.293	0.463	0.527
PhD	0.098	0.955	0.919
country_2			
Colombia	0.028	0.366	0.940
Czech Republic	0.094	0.679	0.890
France	0.320	0.675	0.636
Germany	-0.078	0.565	0.890
Ireland	-0.619	0.585	0.291
Italy	-1.550	0.595	0.010
Netherlands	-0.091	0.374	0.808
Panama	-0.179	0.782	0.819
Philippines	-0.222	1.006	0.826
Portugal	-0.278	0.385	0.472
Qatar	0.324	0.606	0.593
Spain	-0.105	0.476	0.826
UK	-0.661	0.375	0.079
USA	-0.153	0.408	0.708
_cons	2.975	0.534	0.000
Observations		352	
R-Square		0.524	

B. Effect of IV on Mediator

reliability	Coefficient	std. err.	P>|t|
seeming_status_binary			
Authentic	-0.219	0.109	0.046
sentiment_binary			
Positive	0.553	0.109	0.000
age	0.020	0.007	0.007
gender_2			
Male	-0.088	0.120	0.461
education_2			
High-School	0.240	0.168	0.153
Master	-0.294	0.151	0.051
Other	-0.421	0.213	0.049
PhD	0.329	0.721	0.648
country_2			
Colombia	0.374	0.359	0.298
Czech Republic	0.380	0.640	0.553
France	1.241	0.812	0.127
Germany	0.644	0.496	0.195
Ireland	-1.212	0.558	0.031
Italy	0.168	0.977	0.863
Netherlands	0.056	0.374	0.880
Panama	-0.716	0.369	0.053
Philippines	1.852	0.339	0.000
Portugal	-0.379	0.368	0.305
Qatar	0.958	0.829	0.249
Spain	0.343	0.420	0.415
UK	0.264	0.374	0.480
USA	0.347	0.386	0.370
_cons	4.008	0.467	0.000
Observations		352	
R-Square		0.1875	

C. Effect of IV and Mediator on DV

purchase_intention	Coefficient	std. err.	P>|t|
reliability	0.179	0.072	0.014
seeming_status_binary			
Authentic	0.336	0.125	0.008
sentiment_binary			
Positive	2.216	0.139	0.000
age	-0.012	0.010	0.237
gender_2			
Male	0.235	0.138	0.091
education_2			
High-School	-0.357	0.183	0.052
Master	0.408	0.183	0.027
Other	-0.218	0.464	0.639
PhD	0.039	0.956	0.968
country_2			
Colombia	-0.039	0.376	0.917
Czech Republic	0.026	0.714	0.971
France	0.099	0.701	0.888
Germany	-0.193	0.560	0.730
Ireland	-0.403	0.532	0.450
Italy	-1.580	0.557	0.005
Netherlands	-0.101	0.386	0.793
Panama	-0.051	0.779	0.948
Philippines	-0.553	1.014	0.586
Portugal	-0.210	0.394	0.595
Qatar	0.153	0.662	0.818
Spain	-0.166	0.483	0.732
UK	-0.708	0.376	0.061
USA	-0.215	0.417	0.607
_cons	2.259	0.610	0.000
Observations		352	
R-Square		0.5352	

7.6 Moderation Test – Study 2 on Reliability

reliability	Coefficient	std. err.	P>|t|
seeming_status_binary			
Authentic	0.575	0.356	0.107
sentiment_binary			
Positive	1.348	0.334	0.000
interaction1	-0.530	0.217	0.015
age	0.020	0.008	0.008
gender_2			
Male	-0.088	0.119	0.458
education_2			
High-School	0.240	0.165	0.146
Master	-0.294	0.152	0.053
Other	-0.421	0.208	0.044
PhD	0.329	0.669	0.623
country_2			
Colombia	0.374	0.333	0.262
Czech Republic	0.380	0.633	0.549
France	1.241	0.780	0.113
Germany	0.644	0.480	0.180
Ireland	-1.212	0.573	0.035
Italy	0.168	0.936	0.857
Netherlands	0.056	0.350	0.872
Panama	-0.716	0.350	0.041
Philippines	1.852	0.311	0.000
Portugal	-0.379	0.335	0.258
Qatar	0.958	0.848	0.260
Spain	0.343	0.399	0.391
UK	0.264	0.349	0.450
USA	0.347	0.359	0.335
_cons	4.405	0.475	0.000
Observations		352	
R-Square		0.202	

7.7 Moderation Test – Study 2 on Purchase Intention

purchase_intention	Coefficient	std. err.	P>|t|
reliability	0.158	0.071	0.026
seeming_status_binary			
Authentic	1.249	0.424	0.003
sentiment_binary			
Positive	3.145	0.388	0.000
interaction1	-0.612	0.251	0.015
age	-0.011	0.010	0.255
gender_2			
Male	0.233	0.137	0.091
education_2			
High-School	-0.352	0.180	0.052
Master	0.402	0.183	0.028
Other	-0.227	0.432	0.600
PhD	0.046	0.967	0.962
country_2			
Colombia	-0.031	0.415	0.940
Czech Republic	0.034	0.678	0.960
France	0.124	0.772	0.872
Germany	-0.180	0.587	0.760
Ireland	-0.428	0.602	0.478
Italy	-1.577	0.550	0.004
Netherlands	-0.100	0.423	0.813
Panama	-0.065	0.825	0.937
Philippines	-0.514	1.034	0.619
Portugal	-0.218	0.431	0.614
Qatar	0.173	0.629	0.784
Spain	-0.159	0.519	0.760
UK	-0.702	0.416	0.093
USA	-0.208	0.453	0.646
_cons	2.801	0.658	0.000
Observations		352	
R-Square		0.5434	