The effect of (negative) emotions on product evaluations in times of a pandemic

A study on emotions within Amazon product reviews before and during the rapid spread of COVID-19.

Master's Thesis

Merith Jager, 565902

Thesis supervisor: dr. C.S. Bellet

Second assessor: dr. A. Alfons

Erasmus School of Economics

MSc. Data Science and Marketing Analytics

19-07-2021

Executive summary

The COVID-19 pandemic has had a tremendous impact on the lifestyle and mental states of consumers over the last one and a half years. The virus caused millions of people around the globe to be ill and in need of health care. Since March 2020, worldwide restrictions have been introduced which impacted consumers' emotions and caused changes in shopping behaviour as hoarding and impulse buying increased. To discover the direct and indirect impacts of COVID-19 on product evaluations, reviews of product categories that were popular during the pandemic have been scraped from Amazon.com and sentiments within the review text have been examined. Previous literature found that consumers are more likely to exert negative emotions in their product reviews when experiencing emotions of fear and sadness. The information shared by customers provides insights into the drivers of consumer buying behaviour. The main predictive analysis of this study has been preceded by a between-research design survey that contained 92 respondents and aimed to quantify the relation between COVID-19, emotionality, purchase intentions and willingness to write a review. No significant negative effect of fear of COVID-19 on emotionality has been found in this pre-test, but individuals feeling unhappy are more likely to write a negative review. In the pre-test, emotionality has been directly measured by asking respondents questions about their positive and negative feelings. In the main study, emotionality has been measured as emotional intensity of the review. The main study focused on distilling sentiments from review text for both utilitarian and hedonic product alternatives in two timeframes. In 2020, during COVID-19, and 2019, before COVID-19 existed. It emphasises the negative emotions, emotional intensity, and review valence of the review. These features have been used as predictors in two machine learning algorithms: a binary logistic regression and a Random Forest. The product rating scale from 1 to 5 was split into <5 and 5 and was the outcome variable in the main study. The binary logistic regression, including all features as predictors, resulted in the most accurate predictions when running the trained model over a hold-out test sample. Therefore, this model has been used for feature importance determination. It is found that sadness is more present in the reviews and more important in predicting the product rating in times of the pandemic compared to regular times. Also, significantly more emotional words have been used in reviews in 2020 compared to 2019 and ratings were lower in 2020, in line with expectations. For hedonic product alternatives, the number of emotional words used was on average 1.4 times larger than for utilitarian alternatives. Multiple studies investigating hedonic products concluded they result in more emotional attachment due to the appealing attributes and luxurious appearance. The most important positive predictor for the binary variable rating is review valence, measured with a polarity score. It has a relatively larger average contribution in the best performing model in 2020 compared to 2019, which indicates that valence of reviews

determines the rating more firmly in 2020. The review valence has also been found to mediate the relationship between emotional intensity of the review text and the binary product rating in 2020. Additionally, this research showed that the price category of the product and the month in which the review was written also influence the product rating. Future research is recommended with real-life measurements on sadness as a result of COVID-19 to conclude its impact on online review communities, as in the pre-test, the focus relied on simulated fear. To summarize, this study highlights the importance of the impact of emotions on product evaluations in times of a pandemic, aiming to uncover the influences of COVID-19 on negative emotional states. The findings of this study are insightful for product owners and marketing managers since sentiment analyses with polarity scores give a strong indication of the direction of the rating. If an algorithm is built within the review page, product owners can be directly aware of negativity among consumers by distilling their sentiments. It provides insights into what customers are negative about. Emotionality is not necessarily damageable and can even increase brand loyalty of consumers. However, negative reviews are more extensively read by consumers, and they negatively influence product attitudes. Therefore, they need to be diminished where possible. The models as trained in this study can be used to make accurate rating predictions. Brand loyal customers can be filtered, and marketing managers can positively influence negative customers with their marketing communication.

Table of contents

1. Introduction

2020 has been a year of social distancing and self-isolation. Many countries around the world encouraged individuals to stay at home to stop the spread of COVID-19. This is the disease caused by the coronavirus. The World Health Organization first learned of this new virus on December 31, 2019 in Wuhan, China. People infected can get very ill and in need of health care. The rapid spread of the virus resulted in restrictions in affected countries, and the health crisis has been labelled a pandemic in the beginning of 2020. This ongoing pandemic caused great lifestyle changes and awakened strong negative emotional responses such as fear and sadness (Ge et al., 2020). Consequently, consumer behaviour changed due to the so-called lockdown (Sheth, 2020).

Within a lockdown, stores are closed for safety reasons. Shopping behaviour must shift towards online channels. Behavioural researchers discovered changes in shopping patterns over 2020 (Addo et al., 2020; Grashuis et al., 2020; Laato et al., 2020; Sheth, 2020), and the impact of the pandemic on emotions has been investigated (Pedrosa et al., 2020). However, little research has been done on the indirect effects of COVID-19 on product evaluations in the e-commerce industry. This study dives into the emotions expressed in product reviews and ratings to discover differences in 2020, during the pandemic, compared to 2019 when COVID-19 did not yet exist. Natural language processing has been applied with sentiment analysis. Review texts were unnested, and the separate tokens have been labelled based on two pre-determined emotion lexicons. Subsequently, they have been used for predictive analysis.

Reviews are often a source upon which consumers base their purchase decisions (Yoo & Gretzel, 2008). Positive reviews that spread rapidly can lead to organic advertisement for the product online. This can lead to growing brand recognition and increase sales, while one negative online conversation can cause costly damage. Consumers perceive reviews as more valuable than marketing efforts from a company (Gupta & Harris, 2010). Review content can either be cognitive or emotional (Lee & Koo, 2012), while individuals can share information about the product itself or their emotions and experiences related to the product. Dynamic content within online reviews influences product attitudes (Zablocki et al., 2019). Due to negative emotions possibly caused by COVID-19, reviews can be more negatively loaded, which is damageable for the sales and brand of the products. The main research question of this study is:

“How do emotions within reviews influence product evaluations in times of a pandemic?”

A pre-test has been performed to qualify the first assumptions derived from previous literature about the direct effects of COVID-19. Next, a main study that divides reviews into two timeframes and two product types has been performed to draw conclusions on the indirect effect of COVID-19. The product types considered are utilitarian and hedonic. This division provides more detailed insights into the impacts of emotions on different product categories and attributes. It is concluded in many studies that emotions have a more substantial influence on attitudes towards hedonic products (Arnold & Reynolds, 2003; Dahr & Wertenbroch, 2000; Lu et al., 2016; Ren & Nickerson, 2019). After an extensive theoretical framework, the pre-test results are discussed. Not only is it highlighted how COVID-19 affects (negative) emotions, purchase intentions and willingness to write a review, the relations between these four variables are also evaluated. Afterwards, the results of the main study are assessed. Due to the pandemic, there has been a shift in popularity between product categories, but has there also been a change in product evaluations?

Sentiment analysis has been carried out on review content of utilitarian and hedonic products from reviewers located in the United States. The reviews have been retrieved from the website of Amazon.com. The United States generally score high on overall net optimism (see Appendix 1), but this dropped in April and May 2020 (McKinsey & Company, 2020). This research attempts to discover what motivates consumers to write reviews in uncertain and fearful times.

Purchase behaviour is a widely recognized topic in marketing research. Many point out that planned or impulse buying behaviour is influenced by utilitarian and hedonic motivations (Leverin & Liljander, 2004; Yu & Bastin, 2010; Kronrod & Danziger). In this research, the product type is considered a moderator to predict the relation between emotions and product evaluations expressed in a numeric rating. Utilitarian and hedonic products can be separated through their attributes and goal of use. Utilitarian products are described as practical, evaluated purely based on utility. Hedonic products are fun and playful (Babin et al., 1994; Hirschman & Holbrook, 1982).

Impulsive buying behaviour increased due to COVID-19. It is needful to understand the impact of this behaviour on consumers' product evaluation. Marketeers and e-commerce managers should understand the consequences of this behaviour and the direct and indirect influences on their brands and products. The following conceptual framework is drawn within the scope of this research to illustrate the relations expected to be found.

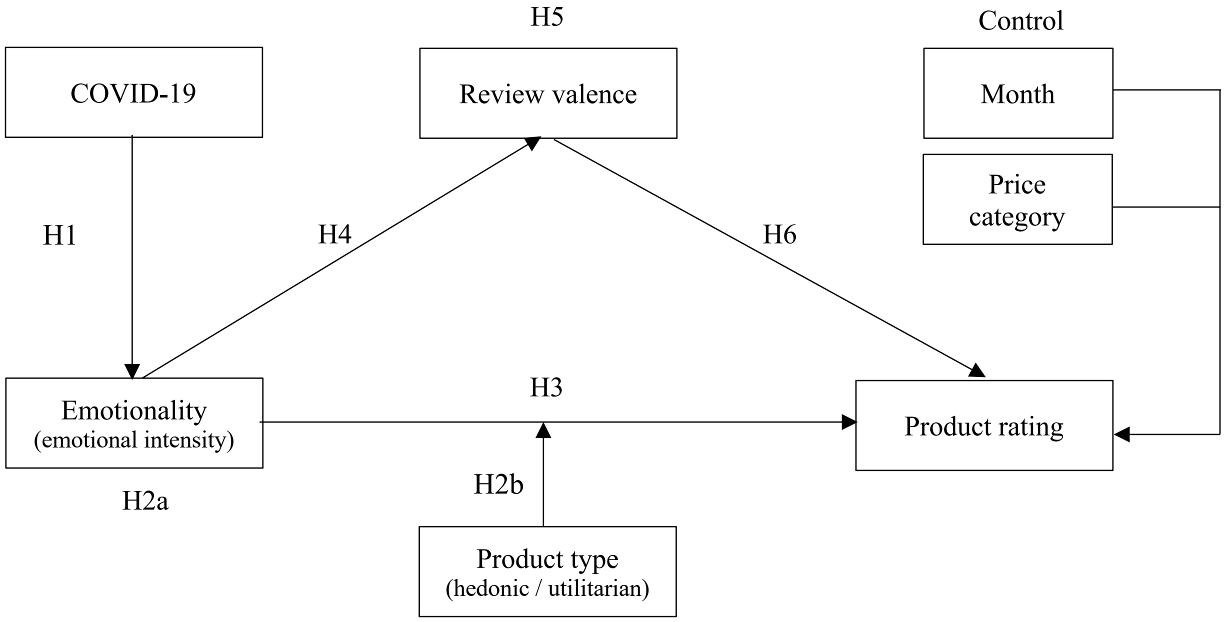


Figure 1. *Conceptual framework*

Five different classifiers are trained and used for prediction: four binary logistic regression models and one Random Forest model. They are trained on an estimation sample and evaluated on a test sample. A rating is a numerical starred grade on a scale from 1 to 5 provided by a reviewer based on their post-purchase evaluation. The rating variable in this research is transformed into a binary variable with two levels: <5 and 5. It is tested which classifier is best suited for predicting this binary variable of product rating.

The final statistical results and implications may apply to online review communities in general and are therefore valuable for managers to consider. COVID-19 is a very recent, impactful event and the out-turns of this pandemic are not yet visible in every industry. However, in the e-commerce industry, results started to show quickly while effects on purchase behaviour were impulsive and direct. Online review platforms are one of the most important sources of information about opinions on a company's product or service. By improving the analysis on sentiment within reviews, underlying causes of negative reviews can directly be uncovered, giving the product owners insights into what drives consumers' negativity. This can then be overcome by responding fast during uncertain and rapidly developing events as a pandemic, where the aim should be to come close to the motivations of consumers.

Overall, customer satisfaction, loyalty and retention should be stabilized or increased in times of a pandemic. The lockdown has to be seen as an opportunity, not a threat. Consumers are buying more online than ever before. Product managers now have their customers all in one place and can influence them by well-advised marketing communication. This can distinguish their products from alternatives to a gain competitive advantage. Proactively searching for the most useful, positive reviews and sharing this with (potential) customers improves the pre-purchase searching process for those customers.

Former studies in 2020 and the beginning of 2021 focused on the direct impacts of the pandemic on sentiment, but no translation on e-commerce and product evaluations has been carried out (Racherla & Friske, 2012). This study contributes to this literature by analysing sentiment in times of the pandemic and construing this in a study of product evaluations for both utilitarian and hedonic products. Product managers and marketers should act upon the insights gained in this study and focus on the impacts of emotions on review text.

2. Literature Review

In this chapter, a literature review is delineated on all the variables included in the research. Firstly, insights into shopping behaviour in times of a pandemic are generated. Secondly, utilitarian and hedonic product types are discussed. Next, electronic word of mouth is explained, whereafter emotional states are evaluated by explaining their origin and their relation to review extremity. Lastly, the dependent variable of the research, product rating, is discussed by looking back on previous findings in the literature.

2.1 Shopping behaviour in times of a pandemic

In January 2020, the first human being was diagnosed with coronavirus in Wuhan, China. The unexpected fast expansion of the virus caused a worldwide pandemic. A pandemic is defined as *'an epidemic occurring worldwide, or over an extensive area, crossing international boundaries and usually affecting a large number of people'* (Kelly, H., 2011). Overall production slowed down, and the growth rate of the global GDP declined. Therefore, the pandemic has had an enormous impact on the global economy. Besides, production and consumers' demand changed with higher product prices and weak consumption. Valaskova et al., (2015) stated that consumer behaviour is generally a constant decision-making process containing searching, purchasing, using, evaluating, and disposing of products and services. However, every consumer is unique in their perceptions about situations with large effects, such as an economic crisis or a pandemic (Amalia et al., 2012). Emotions of fear and sadness were experienced more often, which has been proven to increase the amount of money people spend to purchase items (Garg & Lerner, 2013). Moreover, Addo et al., (2020) state that fear during the pandemic is associated with dynamics in online purchases related to COVID-19. It promotes social presence, and consumers seek affection, acceptance and social information. Due to the pandemic being a very recent and abrupt event, this study aims to surpass current literature and investigate the extent to which COVID-19 influences emotions, purchase behaviour and, in particular, product evaluation.

Other immediate impacts of the coronavirus on consumption behaviour, as stated by Sheth (2020), are summarized in Figure 1. Consumers have been hoarding, which is marked by an overwhelming desire to collect items and an inability to discard things that may seem useless. Besides, embracing digital technology is also considered as an immediate impact of COVID-19. According to a recent survey of adults in the U.S., 37% of survey respondents considered shifting to online shopping due to COVID-19, and 73% exhibited loyalty for intending to continue (McKinsey & Company, 2020;

Morning Consult, 2020). Besides, Sheth (2020) considered pent-up demand as an immediate effect of COVID-19.

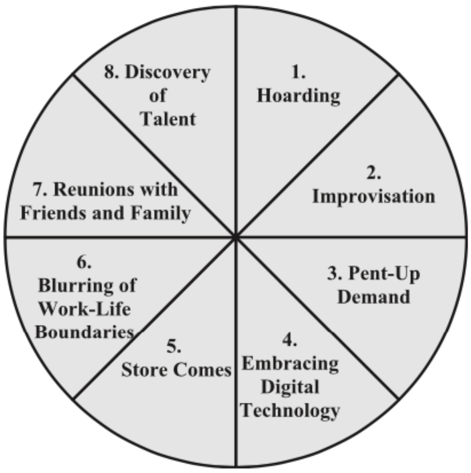


Figure 2. *Immediate Impact of Covid-19 on consumption behaviour (Sheth, 2020)*

The first outbreaks of the virus started to enhance panic buying. Especially Australia (BBC, 2020) and Singapore (INSEAD, 2020) coped with extensive panic buying without a particular indication of stocks shortage. The abnormal high demands led to increasing prices and stock-outs. Products trendy were food items and household supplies. Panic buying increases consumer anxiety and worsens it (Allon & Bassamboo, 2011). Laato et al., (2020) also found a positive link between self-isolation and impulsive, unusual buying behaviour. They call this the stimulus-organism-response (SOR). They derive the cause of this being cyberchondria: the unfounded escalation of concerns about common symptoms based on online search results.

Li (2015) confirms this and found that impulsive buying directly affects mixed emotion responses and affects consumers' post-purchase satisfaction. Previous research has pointed out that planned or impulsive buying behaviour is not only influencing post-purchase satisfaction but is on itself greatly influenced by emotional motivations (Leverin & Lijander, 2004; Yu & Bastin, 2010). Boutsouki (2019) has focused on perceptions and assumptions of economic crises and their negative impact on purchases, but not with actual events. Differences in underlying emotional states in times of a pandemic can be measured by comparing reviews of the same product categories during COVID-19, in 2020, and before COVID, in 2019. The following hypothesis is formulated:

H1: COVID-19 increases negative emotions experienced by consumers

2.1.1 Product category demands during COVID-19

Due to the pandemic, consumer expenditure within specific product categories has changed (Grashuis et al., 2020; J.P. Morgan, 2020). Non-essential items were less prevalent at the beginning of the lockdown. Hoarding caused a significant shift in popularity. Especially products in health, coffee, soap, and fitness equipment categories increased exponentially in sales. Besides, sales of alcoholic drinks and other beverages also increased immensely due to the experience of negative emotions (Pedrosa et al., 2020). Health-conscious consumers loaded up on vitamins and supplements throughout the year. A climb of 50% has been seen in the first half of the year after the virus outbreak. The product category that eventually went through the most prominent growth has been household cleaners and soap. The United States started spending more on groceries in April 2020, but also on home entertainment.

2.2 The distinction between hedonic and utilitarian products

As more is explained about the impact of COVID-19 on shopping behaviour, this section zooms in on different types of products. Products or services can be categorized as search or experience products and utilitarian or hedonic based on the goal of use. A product is utilitarian when it is more rational and used to accomplish a goal, and a product is hedonic when used to achieve enjoyment (Babin et al., 1994; Hirschman & Holbrook, 1982). Search goods contain products that can be evaluated before used, while experience goods can only be evaluated after usage.

Utilitarian products are evaluated based on their characteristics, such as usefulness and practicality, whereas hedonic products are evaluated base on enjoyability, excitement and fun (Voss et al., 2003). It is argued that for utilitarian products, the product's functionalities affect attitudes, and for hedonic products, the product's emotions and how it makes the consumer feel is affecting attitudes. A product can have both utilitarian and hedonic attributes. Voss et al. (2003) elaborate on this with the example of athletic shoes. They are functional and enhance performance but choosing a likeable brand and appealing look are hedonic attributes of the shoe. In Table 1, examples are displayed of utilitarian and hedonic products.

	Utilitarian	**Hedonic**
Search	Printer	Stereo installation
Experience	Car insurance	Cruise trip

Table 1. *Hedonic and utilitarian products (Voss et al., 2003)*

Lu, Liu and Fang (2016) also reported a list of utilitarian and hedonic products in different product categories. For example, a documentary contains primarily utilitarian attributes, whereas a comedy is categorized as hedonic. Products with a lot of emotionally loaded reviews are likely fall into a hedonic goal of use. Reviews on utilitarian products contain a more informative value. Moore (2015) goes one step further in explaining the differences between these product types and introduces two explanation types for the evaluation of utilitarian and hedonic products in relation to reviews. She states that utilitarian reviews are 'action' explanations while they are cognitive, such as "I chose this product". On the contrary, hedonic reviews are 'reaction' explanations coming from the writer's feelings, such as "I love this product". Dhar & Wertenbroch (2000) explained this in more detail in their study. They concluded that while hedonic goods are unique and irreplaceable due to their luxurious value and appealing aspects, consumers develop emotional attachments to them over time and are more reluctant to forfeit them.

This study focuses on search goods for both utilitarian and hedonic product types. Due to the expected impact of the product type on emotionality and the use of emotional wordings in the review text, product type has been included as a moderator in the main study to examine whether it strengthens the relation between the emotional intensity and the rating. The following hypotheses are derived:

H2a: The number of emotional wordings in reviews is higher in times of a pandemic and higher for hedonic products compared to utilitarian products

H2b: The product type strengthens the relation between emotional intensity of the review text, and product rating

2.3 Electronic Word of Mouth

In the purchase decision-making process, consumers tend to rely on information from other consumers. Before online communities for sharing opinions existed, consumers turned to their relatives or professional acquaintances for information about a product, known as Word-Of-Mouth (Arndt, 1967). It is defined as the transfer of information from one individual to another. The internet has caused a revolution in possible ways of communicating via social media platforms. Information and opinions can be retrieved from any consumer at any time. Opinions are expressed in textual reviews, combined with numerical ratings. The rating provides a content indication about the review text, and together, they form a product evaluation. The review text can contain reasons for buying the product, figurative wordings or feelings towards a product. Reviews can save time for consumers while providing insights into product details they no longer have to search for themselves. Online communities where reviews are available are a segment of Electronic Word-Of-Mouth (eWOM). This is defined as *'any positive or negative statement made by potential, actual or former customers about a product, which is made available to a multitude of people and institutions via the Internet'* (Hennig-Thurau et al., 2004).

To compensate for the inability to communicate non-verbally online, reviewers express their opinions with emotional wording (Chen, 2020). Word-of-Mouth activation is the term used to denote an individual's decision to engage in Word-of-Mouth communication. This decision is personal and mainly goal-oriented. If there is no purpose, people are not motivated to communicate with each other. Previous findings show that the decision to engage in eWOM depends on multiple factors. The popularity of a product influences the decision to engage in eWOM (Zhu & Zhang, 2010), together with the consumer's level of satisfaction. Consumers who are very satisfied or very dissatisfied with a product are more likely to post their reviews (Hu et al., 2009). The identity and geographical location of the consumer also plays a sufficient part (Forman et al., 2008). Lastly, previously posted reviews also influence a consumer's review decision (Moe & Schweidel, 2012). The more positive the previous reviews, the more likely consumers are to engage in eWOM. One of the most prominent examples of a review platform that combines both review text and ratings is Amazon.com (Floh et al., 2013, Mudambi & Schuff, 2010).

Another term widely used to indicate the impact of a review on other consumers is review helpfulness (Huang et al., 2015). Product reviews that contribute to the purchase decision making process by including relevant information that helps other customers understand the quality and performance of a product are seen as helpful. Huang et al. (2015) also stated that if one's previously posted reviews were voted as helpful, upcoming reviews are also seen as helpful. Amazon, one of the largest online retailers in the world, has a very advanced review system. Top reviewers are designated and featured on top of review pages. The ability of a website to provide helpful information on product characteristics positively influences consumers' product pre- or post-purchase evaluations (Jiang & Benbasat, 2007).

2.4 The definition of emotionality

Emotions are multicomponent response tendencies that unfold over relatively short periods (Frederickson, 2001). Shaver et al., (1987) divide emotions into six basic emotions: joy, love, surprise, anger, sadness and fear. Plutchick (1978) adds anticipation and trust. The emotion of fear turns out to involve the most mindful processing. Fear rises in combination with high uncertainty, and Yin et al. (2014) characterize this as one of the two main characterizations of discrete emotions within reviews. As mentioned earlier, due to COVID-19, this is a very pervasive emotional state for consumers in 2020. The phenomenon 'negativity bias' refers to the notion that situations of a more negative nature have a greater effect on psychological states of mind than neutral or positive cases (Baumeister et al., 2001; Rozin & Royzman, 2001). Emotionality is the quality or state of being emotional, which is the observable component of emotion.

Sentiment reflects the deeper psychological state of the holder (Hovy, 2015) and can be defined as an attitude, thought or judgment prompted by feeling (Fang & Zhan, 2015). The emotional use of words within a review can be referred to as sentiment. It is considered worthwhile to investigate the drivers of emotional wordings within reviews and different emotion types, especially in times of a pandemic. Consumers who have an emotional connection with a brand have a 306% higher lifetime value (Morgan, 2019). Active processing of emotions is essential when writing an online review. Negative reviews are not necessarily derived from negative emotions, while they can also be stressed from product weaknesses or problems. The review text will then be more informative (Zablocki et al., 2019).

2.4.1 Experiencing emotions in times of a pandemic

According to previous research, COVID-19 negatively influences emotions. The fear of insecurity, uncertainty, loneliness and great changes in lifestyle have awakened strong emotional responses (Ge et al., 2020). When preferences are subject to public evaluation, the emotion of loneliness causes consumers to conform to the norm because they fear being evaluated negatively (Wang et al., 2012). Emotions of fear and sadness are common consequences of the COVID-19 pandemic (Addo et al., 2020). Fear appeals are persuasive messages designed to communicate facts or scare individuals by resenting terrible outcomes of neglecting a specific caution. Marketers often use this to persuade customers. An increase in fear causes an increase in compliant behaviour. Fear is an adaptive mechanism for humans to cope with threats. It can cause defensive behaviour, and it positively associates with depression and anxiety. Incidental emotions such as fear and sadness have been found to influence numerous aspects of judgment, and decision making, such as price spent and risk-seeking (Johnson & Tversky, 1983; Lerner & Keltner, 2001).

A rise has been seen in mental disorders (Brooks et al., 2020; Holmes et al., 2020). Pedrosa et al. (2020) detected other psychological consequences of stress, anxiety, depression, and alcohol addiction. As would be expected, negative emotions should trigger a negative valuation of products. However, in comparison to anger and fear, sadness has also been found to trigger positive valuation of products measured by willingness to pay (Lerner et al., 2004). Sadness enhances the amount people spend to purchase items (Garg & Lerner, 2013). Lastly, in the research of Yin et al., (2014), anger is shown to be mindless and heuristically processed with little thought. This study focuses on gathering underlying emotions, measured and described in two ways – their valence (polarity) and the negative emotion type (fear, sadness, anger). It is presumed that reviewers who intensely experience emotions exert a more extreme rating. The main study investigates whether emotions such as fear, sadness or anger, also lower the overall product rating. These are the emotions considered as negative and used as predictors in the main study.

H3: Reviews with expressed negative emotions (fear, sadness and anger) are associated with a lower overall product rating. This relation is stronger in times of a pandemic

2.5 Product evaluation: the product rating

Previous research has focused on the influence of emotions within reviews on product attitudes (Kim & Gupta, 2012; Peng et al., 2014; Moore, 2015; Zablocki et al., 2019). Differences have been reported between positive and negative emotional wording in the review text. Peng et al. (2014) define emotional intensity within a review as a percentage of emotional content within the review text. They also stated emotional intensity could be interpreted as lying or fraud by readers. An overload of positive emotions within a review is likely to be perceived as not helpful due to suspicion of the reviewer's honesty compared to negative emotions. Sellers can manipulate them to improve sales. Moderate positive emotional use of words can increase trustworthiness and authenticity due to the writing style reflecting real customer experiences. Besides, negative emotional expressions in a single negative review decrease its informative value (Kim & Gupta, 2012). Potential consumers might consider the reviewer irrational and do not let the review affect their attitude. The direction of a review text is positive, negative, or neutral, also known as review valence. Negatively loaded reviews are very likely to be more emotionally loaded compared to positive reviews. This is investigated in more detail, and the following hypothesis is stated:

H4: Reviews with a negative valence contain a higher overall emotional intensity compared to neutral or positive valence

In most online review platforms, it is mandatory also to provide a rating score. The rating indicates how satisfied the reviewer is with their (post) product experience. Reviews with a negative rating (1 or 2) tend to be perceived as more valuable by consumers (Racherla & Friske, 2012). Content of negatively rated products is more extensively read than positive review content because of curiosity about underlying arguments. If this includes negativity about the brand or product, this can be damageable. Ratings of 3-stars are considered neutral, whereas 2 and 4 are moderately extreme and can be labelled with a “1-score” extremity. Ratings of 1 and 5 are extreme and get a “2-score” label for extremity. Extremely rated reviews are positively associated with emotional intensity (Mudambi & Schuff, 2010). Extreme ratings lead to more figurative language in review texts (Mudambi & Schuff, 2010), while explanatory language is found to cause more moderate ratings (Moore, 2012). Besides, when the review is written in the reader’s native language, the intensity of the effect of the emotional wording is higher (Langhe et al., 2011). Another interesting finding on language is that positive words appear more frequently in the English language compared to negative words (Rozin et al., 2010).

Emotional states can activate review writing which plays a part in information spread from one person to another. The valence of the reviews is essential in the mediating role of emotional intensity on the product rating. According to the two-dimensional theory of affect, reviews with similar valences can contain different levels of emotionality (Russell, 1979, 1980, 2003). This is investigated in more detail and the following hypothesis is formulated based on these findings:

H5: Review valence positively mediates the effect of the emotional intensity of the review text on the product rating.

Previous literature state reviews influence product attitudes (Lee & Koo, 2012; Zablocki et al., 2019) and purchase intentions (Zhang et al., 2014). A product attitude is an individual’s belief that using a product will lead to consequences, evaluated in a good-bad dimension. They influence product choice and evaluation (Fazio & Petty, 2008), thus determining if the reader of a review will buy the product. Therefore, it is important to discover the indirect influences of emotions on review text and rating gain managerial insights in how to enhance positive word-of-mouth within reviews to overcome a decrease in sales. Attitudes are composed of cognitive and affective components. Cognitive about knowledge consumers have about products, which might include product-related and non-product-related associations (Keller, 1993). Affective components cover emotional associations.

Due to COVID-19, reviews in product categories where purchase behaviour changed due to the pandemic might contain differences in valence which could in turn influence product attitudes. Therefore, it is discovered if these changes exist and if they impact the product rating significantly. It is measured if review valence becomes a more important predictor for the product rating. The final hypothesis is formulated:

H6: Review valence is a more important predictor for the product rating in times of a pandemic compared to regular times

2.6 Control variables

The features that possibly influence the main features within this study are briefly discussed and summarized for understanding. First, the language used in the review text influences product rating (Langhe, 2001). In this study, only reviews written in English are considered. Secondly, the price of the product that is reviewed has an impact on product rating. It is seen as a signal of quality and influences review valence. Consumers might not be very satisfied, but the review valence is still neutral at a low price due to the balance of price-quality (Dodds et al., 1991; Mitra, 1995). Price category is added as a dummy control feature to the main data. Thirdly, the review experience of the reviewer plays a part. A reviewer can become popular in online communities when it writes multiple valuable and helpful reviews. More popular users express a broader range of opinions than less popular reviewers (Goes et al., 2014). They produce more reviews and write more words that are less emotional. They feel like they have the responsibility to provide more valuable and objective information. However, the reviewers' experience is not of influence in the main study while reviews are randomly scraped from the top-rated reviews.

Amazon has a built-in functionality where readers can rate how helpful they consider the review text to be. Therefore, the reviews scraped are helpful and overcome omitted variable bias: the phenomenon where variables that should be included in the final model are not included even though they have a significant influence on the outcome variable. Lastly, a cultural influence exists on emotional experiences. Eastern countries tend to suppress negative, self-centred emotions to maintain social harmony (Burleson, 2003), whereas, in Western countries, self-centred emotions are typical as western cultures encourage individuality. This study only focuses on reviewers located in the United States.

To summarize the theoretical background, a lot has been researched on positive and negative reviews and the reviewers underlying emotions. However, due to the sudden rise of COVID-19 infections, researching how this pandemic influences emotional states, how it impacts review valence and how it influences product evaluations could result in very interesting findings directly applicable for businesses.

Hypotheses

H1a	COVID-19 increases negative emotions experienced by consumers
H2a	The number of emotional words in reviews is higher in times of a pandemic and higher for hedonic products compared to utilitarian products
H2b	The product type strengthens the relation between emotional intensity and the product rating
H3	Reviews with expressed negative emotions (fear, sadness and anger) are associated with a lower overall product rating. This relation is stronger in times of a pandemic
H4	Reviews with a negative valence contain a higher overall emotional intensity compared to neutral or positive valence
H5	Review valence positively mediates the effect of emotional intensity of the review text on the product rating
H6	Review valence is a more important predictor for the product rating in times of a pandemic compared to regular times

3. Methodology

This section explains how the research continues and what methods are used, based on the theoretical background. This research consists of two parts: a pre-test and a main study. A pre-test has been performed to validate assumptions before the main study was executed. The findings of the pre-test are discussed, whereafter the methods and data for the main study are introduced.

3.1 Pre-test

Based on literature, it is assumed that consumers experience more negative emotions in times of a pandemic and translate this into their purchase and review behaviour. These assumptions have been tested to conclude the first hypothesis of this study and enhance the internal validity of the main study. To find an honest and externally valid result, it is essential to gain insights into the current direct impact of COVID-19 on consumers' emotions and which products they are more emotionally involved in.

3.1.1 Pre-test setup

An experimental survey has been set up in a between-subject design, including questions measuring emotions, purchase intentions and willingness to write a review. The first part of the survey contains questions to get insight into the respondents' demographic information together with questions about their shopping behaviour. Secondly, a treatment condition is introduced by measuring fear of COVID-19 among half of the respondents. This condition is used to measure whether asking questions about COVID-19 and making respondents aware of these emotions results in more negative affect in the subsequent questions of the survey. Respondents designated their level of agreement with seven statements about COVID-19 on a five-item Likert scale created by Ahorsu et al. (2020). Answers range from strongly disagree to strongly agree. The minimum score for each question is 1, and the maximum 5. Half of the respondents do not get exposed to the fear of COVID-19 questions.

Next, the current emotional states of all respondents are measured through the Bradburn's Affect Balance Scale (Bradburn, 1969). The scale has been made up out of two components: positive and negative affect, with each five items. The respondent answered "Yes" or "No" and get 1 point for every "Yes". To derive the emotional state, the overall "balance" affect score is calculated by the subtracting negative affect score from the positive affect score. When someone is experiencing primarily negative emotions, the score will likely be lower than 0. The ABS scale was founded in

1969 but is still widely used. Respondents were first asked to imagine how they felt in the first lockdown in 2020 to measure emotions in the same time frame as the main study.

Afterwards, respondents' purchase intentions for a utilitarian and hedonic chair alternative were measured. Intentions to buy were discovered by asking questions from a study by Lu et al., (2014). This study suggested that purchase intention is a consumer's willingness to buy a given product at a specific time in a specific situation. The Theory of Reasoned Action (TRA) (Fishbein & Ajzen, 1999) and the Theory of Planned Behaviour (TPB) (Ajzen, 1985) propose consumer attitudes directly affect behavioural intention and, therefore, will influence purchase intention. In the pre-test, purchase intentions had to be reported for both a desk chair (utilitarian) which is functional and mostly for work purposes, and a lounge chair (hedonic), which has features to relax and is more appealing. After that, their willingness to write a negative or positive review was measured. To measure a real-life situation as much as possible, respondents were asked to imagine they needed a new chair. After measuring purchase intentions, a short situation was displayed. Half of the respondents first had to imagine that the product they received had a broken part. They had to answer how likely they would be to write a negative review. The other half of the respondents, randomly chosen, first had to imagine they were satisfied with the product anyways after using it for a couple of days. They then had been asked how likely they are to write a positive review.

All variable items should have a Cronbach's alpha larger than .70 to cover internal validity (DeVellis, 2003). Cronbach's alpha is a measure of internal consistency, computed by dividing the product of the number of items *N* and the average covariance between item-pairs $\bar{c}$ by the average variance $\bar{v}$ summed up to the number of items minus 1, multiplied by the average covariance.

$$\alpha = \frac{N \cdot \bar{c}}{\bar{v} + (N-1) \cdot \bar{c}}$$

The Cronbach's alpha of the COVID-19 fear scale is .82, the Bradburn's scale accounts for a Cronbach's alpha of .73, that of the purchase intentions .914 and that of willingness to write a review .93 on average. The survey has been sent out to personal network acquaintances. All survey items are displayed in Appendix 3.

3.1.2 Pre-test results

The survey has been conducted from May 8 till 13, 2021. In total, 125 responses have been recorded. However, 33 respondents did not finish the questionnaire. The answers of the 92 remaining respondents have been used for further analyses. The distribution of male and female respondents was almost even. The age of the respondents ranged from 18 to 64. The experimental survey consisted of five blocks of items with one treatment condition. In Table 2, the demographic information of the respondents is displayed.

Gender	Frequency	Percentage
Male	40	43,5%
Female	52	56,5%

Age	Frequency	Percentage
18-34	83	90,2%
34-64	9	9,8%

Table 2. *Demographic information pre-test*

Correlation is a descriptive statistic that measures the linear relationship between two variables, in this case, survey items. The range of possible values is from -1 to 1 where 1 accounts for a perfect positive correlation. While most items measure responses on a 5-point Likert scale, items can be inspected on statistically significant correlations. If the p-value of the correlation test statistic is smaller than 0.05 it can be concluded that the correlation significantly differs from 0.

In the following sections, all variables included in the pre-test are compared to conclude the assumptions. Survey data is non-normally distributed, while Likert scales are ordinal and not continuous. To statistically test the distributions of the variables, a Shapiro-Wilk's test is performed. This test is based on correlations between data and the corresponding normal scores (Ghasemi & Zahediasl, 2012). The null hypothesis assumes that the sample distribution is normal. All tests were significant, implying that the distribution of the data is significantly different from normal distribution, and normality cannot be assumed.

Mann-Whitney-U tests and Wilcoxon Signed Rank tests are performed to inspect differences between the groups. These tests assume variables to be non-parametric and non-normally distributed. The Mann-Whitney-U test, equivalent to a Wilcoxon Rank Sum test, assumes samples are independent and that two pairs of populations have the same continuous distribution. A Wilcoxon Signed Rank test assumes dependency between samples to compare sample medians against a hypothetical median. The latter is a widely used alternative for the student's t-test which compares the mean of a small sample drawn from a normally distributed populated with an unknown standard deviation. The test statistic W or V is calculated as follows: For each item in a sample of n items, a difference score D_i is obtained between two groups by subtracting one from the other. Positive and negative signs are neglected to generate absolute differences $|D_i|$. Differences of zero are omitted, and $n' < n$ where n' is the actual sample size. Ranks are assigned from 1 to n

to each of the $|D_i|$ such that the smallest absolute difference score gets rank 1 and the largest gets rank *n*. If two or more are equal, they are assigned the average rank of the ranks that would have been assigned to them if these ties did not occur. Next, positive or negative signs are assigned back to the ranks based on their original signs. Finally, the Wilcoxon test statistic is subsequently obtained as the sum of the positive ranks which is compared to the expected rank-sum. Non-zero median difference causes the test statistic to be very large and significant.

3.1.2.1 Shopping behaviour

In Table 3, the Pearson correlation, mean, median and dispersion of the shopping behaviour items are displayed. Pearson correlation provides information on the magnitude of the correlation as well as the direction of the relationship. Standard deviation measures the variation around the average of the item. It indicates the mean's reliability and is known as the root mean square deviation.

Item	ρ	mean	median	st. Dev	min	max
Q3) How often do you shop online?	1	3.152	3	0.709	2	5
Q4) Did your shopping frequency increase due to the COVID-19 pandemic?	0.251**	3.632	4	0.910	2	5

Table 3. *Descriptive statistics shopping behaviour item construct, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$*

These insights specify that, on average, the respondents do shop online. The option 'Never' has not been filled in by any of the respondents. The mean of $\bar{x} = 3.632$ of Q4 indicates that, on average, respondents' shopping frequency increased. This increase is significant ($V = 4013$, $p<0.000$) according to a Wilcoxon Signed Rank test where it was tested if the median of the responses was greater than 2.5, the median of the scale. These two items have a significant positive correlation, designating that the more often a respondent shops online, the more their shopping frequency has increased due to the pandemic.

3.1.2.2 Fear of COVID-19

It is investigated if the treatment condition of fear of COVID-19 affects the emotions, purchase intentions and willingness to write a review of the respondents. Half of the respondents answered questions on their fear for COVID-19 by thinking back to the first lockdown. First, it is examined whether the treatment condition impacts emotions. Emotionality is measured with the ABS scale, whereas afterwards, the overall balance score for every respondent has been calculated. The Cronbach's alphas of the positive and negative affect items have been examined separately. Two items of the negative affect scale have been removed, which resulted in a larger Cronbach's alpha.

Before performing this test, it has been ensured if the data passes the test of homoscedasticity, are the variances homogeneous. The p-value of the F-test turned out to be p=0.931, designating the variances are homogeneous. The total ABS scores did not significantly differ for the treatment condition of COVID-19 fear versus no treatment condition (W = 1089.5, p=0.801) derived from an independent two-sampled Mann-Whitney-U test. While the first hypothesis and main study focus on exploring negative emotions, the analysis for upcoming sections is divided into positive and negative affect scores of the ABS. The treatment conditions did not significantly cause a difference in average positive affect scores (W = 1167, p=0.381) or negative affect scores (W = 1082, p=0.845).

3.1.2.3 Emotionality

The median item-total correlation (corrected for autocorrelation) was 0.49 for the positive affect and 0.57 for the negative affect of the ABS-scale. Reliability coefficients (Cronbach's alpha) were reasonably high at 0.62 for the positive affect scale and 0.64 for the negative affect scale. When investigating single item significance, item Q6_2) *Proud because someone complimented you on something you had done?* positively correlates with the purchase intention of the desk chair on a significant level. This also accounts for Q6_9) *Very unhappy?* for both the desk and lounge chair, indicating that the when the respondent feels unhappy, they are more likely to purchase one of the two chairs. Non-significant correlations with the emotionality items, however, are stronger for the utilitarian desk chair alternative. This is caused by the overall purchase intention being higher for this product. From this survey, it can be concluded that emotionality does not necessarily differ between utilitarian and hedonic products, which is why this is examined in more detail in the main study with sentiment analysis.

Lastly, it is tested what correlations are between emotionality items and willingness to write a negative or positive review. Item Q6.9) *Very unhappy?* of the negative affect scale correlates positively on a significant level with willingness to write a negative review (ρ = 0.10, p=0.05). Respondents that score high on feeling unhappy also are more likely to write a negative review. This is tested in more detail by performing a linear regression with the interaction between a dummy coded feature for the treatment condition (1 = treatment, 0 = no treatment), item Q6.9) *Very unhappy,* and the willingness to write a negative review as outcome variable. The interaction effect turned out to be significantly positive ($\beta_{TreatmentDummy*Q6\_9}$ = 1.038 p = 0.045), indicating that the through the treatment effect, where people were reminded of their fear of COVID-19, the relation of being very unhappy and the willingness to write a negative review is significantly larger compared to both features alone.

Coefficients

	Estimate	Std. error	z-value	p-value
(Intercept)	3.181	0.182	17.484	0.000***
treatment_dummy	-0.3845	0.2503	-1.536	0.128
Q6_9	-0.5360	0.3524	-1.521	0.131
treatment_dummy:Q6_9	1.038	0.5128	2.025	0.045*

Table 4. *Linear regression of interaction between pre-test treatment and Q6_9 feeling very unhappy*

3.1.2.4 Product type: purchase intention

The scale used for measuring purchase intentions contained one item that was reverse coded. This item has been reversed to make it positive and comparable to the other variable items. Purchase intention has been measured for a utilitarian and hedonic product alternative. The Cronbach's alphas of the purchase intentions of the utilitarian and the hedonic product alternative turned out to be 0.79 and 0.91 respectively. For this variable and for willingness to write a review, it has been investigated in more detail if the item measurements of the construct are similar by performing a principal component analysis. This reduces data dimensionality and is a well-suited method to validate if the items load on the same component. In this way, conclusions can be drawn on one variable (component) measured with multiple items, instead of every item separately. Principal components are uncorrelated variables that successively maximise variance and do not require distributional assumptions. A separate data matrix is created with the items of one construct *i* and observations *n,* representing the respondents. These data values define in-dimensional vectors $x_1,\dots x_i$, resulting in a $n\ x\ i$ matrix **X**, whose *j*th column is the vector x_j of observations on the *j*th variable. A linear combination is to be found between the columns (items) in matrix **X** with maximum variance. Such linear combinations can be denoted as $\sum_{j=1}^{i} a_j x_j = \mathbf{Xa}$ where **a** is a vector of constants $a_1,\dots a_i$.

The range of items are standardized to contribute equally to the analysis whereafter the covariance matrix is computed. With 5 items, the covariance matrix is a 5 x 5 symmetrical matrix. In the main diagonal, the variances are projected of each initial item. Next, eigenvectors and eigenvalues are computed to identify the principal components. These are linear algebra concepts. The principal components represent a direction of the data that explains a maximal amount of variance, which are the lines in the dimensional space that capture the most information of the data. The more prominent the relationship between variance and information, the larger the variance carried by the line, the greater the distribution of the data points, the more information it holds. The more variance the first component explains, the more likely the items measure the same construct. The first component for the utilitarian purchase intention explains 55% of the variance, whereas the hedonics' first component explains 73%. Since 50% is an adequate amount of variance explained for consumer behaviour research (Merenda, 1997), the items are assumed to define the same construct. They are summed up and divided by the total number of items for further analysis. Purchase intentions for the desk chair are significantly higher ($W=2007.5$, $p<0.000$), as concluded from a two-sampled Mann-Whitney-U test that combines both product category item blocks. The purchase intentions for both chair alternatives were not different within the two treatment conditions (utilitarian: $W = 1038.5$, $p=0.884$; hedonic: $W = 1079.5$, $p=0.866$).

3.1.2.5 Willingness to write review

For willingness to write a positive or negative review, the variance explained by the first components retrieved from the PCA ranged from 89% to 96%, indicating that all 8 items can be combined into 4 variables by summing the answers and dividing them by 2. It has already been examined if positive or negative affect has stronger correlations with willingness to write a positive or negative review. The mean for willingness to write a negative review is $\bar{x} = 2.970$, while for a positive review this is $\bar{x} = 2.614$ for both product categories. The average rank sums are compared with a Mann-Whitney-U test, and the difference turned out to be significant ($W = 3395$, $p=0.017$). The median of the rank sum of the negative items is significantly larger.

Within product categories, the averages of writing a positive compared to a negative review are significantly different for the desk chair alternative ($V = 796$, $p=0.004$) and the lounge chair alternative ($V = 980$, $p=0.001$). Respondents are generally more likely to write a negative review in both categories.

3.2 Main study

The main study of this research focuses on review data from Amazon.com. Afterwards, sentiment analysis is executed over the review data to draw conclusions on emotions in reviews, test the hypotheses and formulate recommendations concerning retailers and marketers for post-corona strategies.

3.2.1 Data collection

The data used in this research is product review data generated from Amazon.com, written by reviewers located in the United States. Reviews are extracted from the website in two separate time frames using a Google Chrome web scraper. The information scraped in a dataset is the id, the reviewers' profile name, the review rating (on scale of 1-5), the reviewers' location, the date of the review and the review text. Amazon reviews are considered a good source due to the large number of data points. Moreover, real-world observational data rules out artificiality when reviewers are forced to write the review and are not intrinsically motivated (Yin et al., 2014). This would cause a high variation in review length and less emotionality due to a lack of real-world experiences and evaluations. The scraped data is transferred into a csv file which can easily be imported into R to execute multiple analyses.

3.2.2 Machine learning method

In the main study, multiple predictive models have been built. The machine learning method used for extracting sentiment information is known as Natural Language Processing (NLP). This is an area of research that aims to understand and manipulate natural language text or speech for statistical analysis. Text mining is used, which examines large collections of text data to discover patterns and information to help answer the main research question. Text mining algorithms can reveal customer attitudes and sentiments to compare and highlight the opinions of different customers (Jack & Tsai, 2015).

The "bag of words" approach is used to examine words within reviews separately to identify sentiments by labelling the words. Review valence and emotion types are detected for both utilitarian and hedonic product alternatives. Text mining with review data is a suitable method for discovering product attitude changes while it is an analysis performed on already purchased goods. Other methods could include surveys or other interviewing techniques. These methods can only capture willingness to pay or purchase intentions, not actual purchase behaviour. Conclusions could be biased.

3.2.2.1 Sentiment analysis

This section goes into more detail on the first part of the analysis performed on the retrieved data. Sentiment analysis filters an author's emotional intent into separate emotional classes by extracting their emotional intent from the text (Kwartler, 2017). Polarity is the term and method used to create a numeric value for the valence of the review. It is based on subjectivity lexicons that determine if a word has a positive or negative sentiment depending on pre-determined labelled words. The polarity score considers valence shifters: amplifiers and negators. Negators infer the actual polarity, while amplifiers strengthen the review's polarity by increasing its intensity. An example of a negator is "not", whereas an amplifier example is "very". The outcome of the polarity is a numeric value that has a positive or negative sign. If a word is recognised as positive in the lexicon, it gets a value of 1. If elsewhere before an emotional word a negator is found, the sign flips, and the polarity score becomes negative. If an amplifier is present in front of a word, 0.8 is added to the value.

Finally, the sum of the polarity within the review is divided by the square root total number of words. Full review texts are used to discover differences between polarity scores over time and the rating that belongs to the review text.

Two subjectivity lexicons are considered for the analysis, namely: Bing and NRC. They differ in their procedures of labelling words. The Bing lexicon contains 6788 terms and is created by Bing Liu at the University of Illinois, Chicago. This lexicon labels words as positive or negative. The NRC lexicon contains 13,901 words associated with emotions. Eight different emotions are labelled to all the words (anger, fear, trust, surprise, sadness, joy and disgust). The annotations have been manually done by crowdsourcing. For this study, Bing and NRC are used to reveal emotional intensity, negative emotions and review valence the review texts.

3.2.3 Logistic regression and Random Forest

Five different classifiers are trained and used for predictive analysis, using two machine learning algorithms: binary logistic regression and a Random Forest. The data is classified to determine class membership *y'* of an unknown data item *x'* based on data D = $((x_i, y_i), \ldots, (x_n, y_n))$ of data items x_i with known class membership y_i. The attempt to predict class membership can be denoted as *P(y|x)*. Logistic regression is an alternative to ordinal linear regression. Both methods follow the same general principles but are different in parametric and assumptions. Linear regression outcomes are continuous, whereas outcomes of logistic regression are discrete: binary or multinomial. For each of these options, a different logit is used as a measurement. The binomial logit is suitable for two outputs, 0 and 1, and the multinomial logit for more than two outputs. Linear regression fits a straight line to the datapoints. The most straightforward formula to represent linear regression is $y = \beta_0 + \beta_1 x$ where *y* is the outcome variable, β_0 the intercept and β_1 the coefficient. In logistic regression, a sigmoid function is used. In this study, the outcome variable is binary (rating <5 and 5). The logistic function can be rewritten to a logit function with log odds.

$$\frac{p}{1-p} = \exp(\beta_0 + \beta_1 x) \longrightarrow ln\frac{p}{1-p} = \beta_0 + \beta_1 x$$

The line that is aimed to be found maximises the likelihood. This line is the best-fitted line. One unit increase in x changes the expected log odds by β_1. Interpretation of the odds can be done by exponentiating the coefficient, while for every increase in x, the expected odds in favour of *y*=1 increases by exp(β_1). The probability that the outcome is 0 or 1 is discovered. Probabilities range from 0 to 1, and odds range from 0 to positive infinity. Odds are defined as the ratio of the probability of success or failure. Exponentiating, taking the multiplicative inverse of both sides and solving the equation to get $p(y=1)$, results in $p = \frac{exp(z)}{1+exp(z)}$, where p is the probability and $z = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$.

The data is first randomly split into a 70% estimation and 30% test sample for both methods. The test sample is used to evaluate the models. After running the regression model, the confusion matrix is inspected, which designates which outcome variable values are predicted correctly. It demonstrates the asymmetry of positive prediction versus negative prediction. An advantage of logistic regression is that its interpretation is straightforward and relatively easy.

A Random Forest is an advanced ensemble method suitable for classification and regression problems, with a different approach than logistic regression. Ensemble methods combine algorithms for classifying objects to enhance predictive power. The building blocks of a Random Forest are decision trees. Data is split according to a criterion, cross-entropy, also known as log loss. This maximises data separation and results in a tree-like structure. It refers to disorder or uncertainty. A perfect model would have a log loss of zero. At each split, the decrease in entropy is maximised. The formula for binary cross-entropy is $-(y\log(p) + (1 - y)\log(1-p))$, where y is the binary outcome being 0 or 1 and p the proportion of misclassified observation within the subpartition. To ensure trees are uncorrelated, Random Forests make use of bagging: bootstrap aggregation. Random samples, with replacement, are taken from the data that result in multiple trees. A portion of the data is left out, known as the out-of-bag-error (OOB). The model automatically evaluates its predictions by validating them over the OOB sample throughout the forest.

The name Random Forest is derived from the random subset of features considered per split. The rule of thumb is to take the square root of the total input features. It forces even more variation amongst the trees. Every tree delivers an output prediction for outcome variable *y*. The final prediction value is computed by taking the majority vote of the predictions made by each individual tree in the forest. Their ability to limit overfitting without increasing bias makes Random Forest a robust model. Overfitting occurs when a model learns the noise in the train data and performs poorly on new, test data. As Random Forests are trained over different samples of data due to bootstrapping, predictions are likely to be accurate.

Conclusions on feature importance are derived from the model that best predicts the binary outcome variable rating. The metrics used to determine which model best predicts the rating are accuracy, precision, recall and Kappa Statistics. Accuracy is a measure of correct predictions over total predictions. It is based on the difference between the observed values and predicted values. In this case, the training data is compared to the test data. Accuracy may, however, not be the most appropriate metric for this data. The class distribution of the rating can be imbalanced. Therefore, precision and recall are considered for each model. Precision (TP/(TP+FP)) is the proportion of observations that the model classifies as positive, and which are actually positive. Recall (TP/(TP+FN)) is the proportion of actual positives that the model correctly classifies. The Kappa statistic is also interpreted. It compares an observed accuracy with an expected accuracy. The observed accuracy is derived from the number of instances correctly classified. The Kohen Kappa considers random chance, which is less misleading than accuracy (Landis & Koch, 1977). Also, it is uncovered which method for labelling sentiments is best suited for this data.

The McFadden index R^2M, also known as the pseudo-R-squared, is used to compare the predictive power of models with the same predictors in both 2019 and 2020. This metric is comparable to the R-squared used in linear regression problems to get an understanding of the variance explained by multiple models. The metric only has meaning when compared to another pseudo-R-squared of the same type and with the same data. McFadden's pseudo-R-squared is based on the log-likelihood kernels for the intercept-only model (L_{null}) and the model with all estimates (L_c) (McFadden, 1973).

$$R^2_{\text{McFadden}} = 1 - \frac{\log(L_c)}{\log(L_{null})}$$

The aim of a classification model is for $y = 1$ to predict $P(y=1) \approx 1$, and for $y = 0$ predict $P(y=1) \approx 0$. The probability of seeing this come true, and predict right, is 1 for both cases. This means the likelihood value for each observation is close to 1. The log of 1 is 0, so the log-likelihood value $\log(L_c)$ will be close to 0. Hence, R^2_{McFadden} will be close to 1 indicating excellent predictive ability.

3.2.4 Data sources

3.2.4.1 Product categories

For the main study, three product categories are selected based on the statements in the articles of Dhar & Wertenbroch (2000) and Ren & Nickerson (2019). One of the products in every category contains utilitarian attributes (rational, practical) and the alternative hedonic attributes (luxurious, appealing). Moreover, products are selected based on shopping behaviour in times of the pandemic. They are displayed in Table 5.

Amazon category	Product	Utilitarian or hedonic
Home & Kitchen - Furniture	Ergonomic desk chair	Utilitarian superior
	Lounge chair	Hedonic superior
Books	Science book	Utilitarian superior
	Fiction novel	Hedonic superior
Clothing, Shoes & Jewelry	Sport watch	Utilitarian superior
	Fashionable watch	Hedonic superior

Table 5. *Overview of utilitarian and hedonic products used for the main study*

The products are selected based on their description, attributes and the number of reviews they have retrieved in 2019 and 2020. Amazon provides a tremendous number of products in multiple product categories. It is essential to scrape many reviews per product due to concerns with internal validity and variance in the reviews. The average of reviews scraped is 3,884 resulting in a dataset with 23,309 observations and ten features. Out of these observations, 1,129 of the reviewers did not originate from the United States. Therefore, these observations are removed. Moreover, only reviews written no earlier than 2019 or later than the end of 2020 are kept in the final dataset. Finally, only reviews written in English remain. The final dataset contains 6,798 observations.

3.2.4.2 Time frame and location

The research period that the analysis focuses on is divided into two timeframes. These are used to uncover differences in pre-and post-COVID-19. By comparing reviews in 2019 to reviews in 2020, it can be seen what the differences in feature importance and accuracies of predictions are between the periods. Changes are compared over time between these two samples. Each sample has a total period of 11 months. The start date is the 26th of January, as this is the date that U.S. residents got exponentially aware of COVID-19 concluded by Google Trends (2021).

The first coronavirus case in the U.S. has been discovered on the 25th of March. However, the first case had already been reported on the 31st of December 2019 in Wuhan, China. Amazon is originated in the U.S. and has the biggest market in this country in 2020. Therefore, this study focuses on this country. The United States is one of the countries the coronavirus has most largely impacted, with many infections and deaths.

3.2.5 Overview features

3.2.5.1 Dependent variable

Product rating

The dependent variable of this research is the product rating. Conclusions on the rating are drawn based on the sentiment analysis and predictive analysis. The outcome variable in the models is binary, namely ratings of <5 and 5. This causes a classification problem. Predicting rating in the 5 original classes enhances the chance of overfitting due to a large number of 5-star ratings in both 2019 and 2020. The division of the binary outcome variable rating is displayed in Table 6.

	<5	5
2019	1288	2461
2020	1227	1822

Table 6. *Rating division*

3.2.5.2 Independent variables

Negative emotions

Where in the pre-test fear of COVID-19 and negative affect have been measured, the main study divides negative emotions into **fear**, **sadness** and **anger**. They are used separately as independent variables and are derived from the NRC lexicon.

Overall emotionality (emotional intensity)

Secondly, in the main study, emotionality is measured as **emotional intensity** of the review text and used as an independent variable. This is done by summing up the fear, sadness, anger, trust, joy and anticipation emotions of the NRC labelled data, while these are in line with the emotions considered as primary by Shaver et al., (1987) and Plutchik (1980) (see Appendix 2). The sum of the emotional words is divided by the total number of words in the review text and used as a predictor variable, moderating interaction variable with the product type and in relation with the mediating variable review valence.

3.2.5.3 Mediator

Review valence

The mediator in this study is review valence. This is determined by the **polarity** score that calculates a score for the positive or negative loading of the review text, considering valence shifters. To calculate the polarity score, the Bing dictionary has been used. This is different from the NRC lexicon used for negative emotions and emotional intensity, while the Bing dictionary does not specify emotions, just positive or negative valence scores.

3.2.5.4 Moderator

Product type

The moderator in this research is product **type**. The difference in product type is measured by analysing the interaction between product type and emotional intensity on product rating due to the moderating influence of product type. A moderator alters the strength of the relationship between a predictor *x* and outcome variable *y*. The question answered has been for which product type the effect is the strongest.

3.2.5.5 Control variables

A **price** feature is added to control for differences in emotional intensity and rating due to differences in price. This feature ranges from 1 (low price) to 3 (high price). This dummy coded variable covers the effect of low-, medium- or high-priced products to rule out an unnoticed external influence of price on the rating when functioning as a control variable. The chairs are in the high price category, watches in medium and books in low. The influence of the price category is examined in the final analysis to draw conclusions by comparing it to the importance of emotions. Lastly, the feature **month** is included to detect the impact of the month of the year on the rating.

3.2.6 Data preparation

The scraped datasets have been prepared and reduced to obtain only relevant information. The profile name, the title of the review, images and specific product details have been removed from the original dataset. A categorical feature for the utilitarian or hedonic product type has been added. Also, a variable for rating extremity was added for the descriptive analysis. Finally, the polarity and NRC scores were added to the dataset. The following features remain:

Feature	Explanation	Measure
id	ProductID	Product category and type
text	Written content of the review	Textual part of the review, written by the reviewer
type	Product type	Dummy variable utilitarian product or not
rating	Numerical rating of the review	Rating on a scale of 1-5 indicating the overall opinion of the reviewer
rating dummy	Binary coded rating	0 for <5-star ratings and 1 for 5-star ratings
year	The year in which the review is written	Year 2019 or 2020 indicating pre-COVID-19 or during COVID-19 data
month	Month in which the review has been written	Numerical month indicated on a range from 1 to 12

polarity	Positive or negative review loading	Valence score of review text including valence shifters
nr words	Number of words	Total number of words per review summed
nr emotion words	NRC count of emotions	Total number of emotional words fear, sadness, anger, trust, anticipation, joy per review summed
emotional intensity	Ratio of emotions in review text	Number of emotional wordings in review divided by total number of words
rat.extr	Rating extremity	0-score indicates rating of 3 stars, 1-score to 2 or 4 stars, and 2-score referring to a rating of 1 or 5 stars
fear	Emotional wordings of fear	Count of words labelled as fear by NRC lexicon
sadness	Emotional wordings of sadness	Count of words labelled as sadness by NRC lexicon
anger	Emotional wordings of anger	Count of words labelled as anger by NRC lexicon
price_low	Low product price category	Category labelled for books, indicating low prices
price_medium	Medium product price category	Category labelled for watches, indicating medium prices
price_high	High product price category	Category labelled for chairs, indicating high prices

Table 7. *Overview of all features in the main study*

The data is split based on the year the review has been written in, distinguished in reviews from 2019 and 2020. In 2019, in total 3,749 reviews were written, whereas in 2020, in total 3,049. Stop words have been excluded. These are words that are commonly used in the English language, such as “a”, “the” and “is”. They do not carry much useful information do not contain any sentiment. The reviews have not been stemmed, while stemming causes the meaning of the words to change, and this results in the pre-determined lexicons not recognising words that might be emotional. Stemming words is done in many other NLP tasks and removes the end of a word to only remain the informative part. An example is the word “beautiful”, wherewith stemming, the “ful” part is removed.

For logistic regression, multiple assumptions need to be met. Basic assumptions are linearity in the logit for continuous variables, absence of multicollinearity between variables and lack of strongly influential outliers. Variables that are multicollinear are highly linearly related. This weakens the predictions as variances between predictors get high. Multicollinearity occurs when a correlation between two features is higher than 0.7. This is the case for **nr words** and **nr emotion words**. They have not been included together in any of the prediction models. Outliers can be uncovered by checking the distributions of the continuous data. An outlier is a data point that significantly differs from other observations, and it is mostly a result of variability in the measurement. Continuous variables included in the main study are **nr words, nr emotion words, polarity** and the variables measuring negative emotions: **fear, sadness** and **anger**. The Rosner's test can detect multiple outliers at once, instead of just the lowest and highest value. It is an appropriate test when the sample size is large. Also, it is designed to avoid the problem of masking where an outlier that is close to another outlier can remain undetected. The *n* observations are ordered from smallest to largest. The maximum number of outliers expected is specified beforehand and denoted as k, which can range from 1 to 10. The test statistic is then calculated by removing the observation that is farthest away from the mean, and the test statistic is recomputed by:

$$R_{i+1} = \frac{\left|x^{(i)} - \bar{x}^{(i)}\right|}{s^{(i)}}$$

where $\bar{x}^{(i)}$ is the sample mean and $s^{(i)}$ the standard deviation of the data after the most extreme observation *i* is removed. Lastly, $x^{(i)}$ is the observation in the subset of the data that is the most far away from the sample mean. Multiple test statistics $R_{1\ldots}\, R_k$ are computed, with the hypothesis being that *k* outliers exist by comparing R_k to the critical value λ_k for a pre-specified significance level. If none of the tests is significant, it can be concluded there are no outliers in the variable. Outliers in the variables **fear, sadness** and **anger** were removed while there were only 5, 3 and 4, respectively.

4. Data analysis

In this chapter, multiple descriptive analyses are performed to get insights into the main data and differences between the distributions of the timeframes.

4.1 Descriptive statistics

	Mean	Std.Dev.	Median	Min	Max	Kurtosis	Skewness
2019							
rating	4.220	1.29	5	1	5	0.91	-1.51
nr words	13.430	19.71	8	2	360	63.62	6.09
nr emotional words	4.537	6.32	3	0	104	45.27	4.77
emotional intensity	32.22%	28.14%	28.57%	0.000%	100%		
polarity	0.278	0.57	0.277	-1.93	2.30	0.17	-0.05
fear	5.21%	9.42%	0.00%	0.00%	100%		
sadness	2.18%	5.93%	0.00%	0.00%	100%		
anger	1.74%	6.91%	0.00%	0.00%	100%		
rating extremity	1.662	0.89	2	0	2	1.58	-1.60
2020							
rating	3.964	1.44	5	1	5	-0.21	-1.15
nr words	14.80	22.17	30	2	489	128.49	8.30
nr emotional words	4.906	7.16	4	0	132	70.56	6.17
emotionality intensity	31.82%	27.41%	28.58%	0.000%	100%		
polarity	0.288	0.523	0.333	1.145	2,46	0.06	0.03
fear	4.89%	11.23%	0.00%	0.000%	100%		
sadness	2.68%	7.46%	0.00%	0.000%	100%		
anger	2.05%	6.88%	0.00%	0.000%	100%		
rating extremity	1.627	0.62	2	0	2	1.15	-1.54

Table 8. *Descriptive statistics non-categorical variables review data main study*

Overall, the medians and standard deviations are quite similar. Skewness is a measure of symmetry or the lack of symmetry. Kurtosis is a measure of whether the data is heavy-tailed or contains outliers. A high value for kurtosis indicates heavy tails but is dependent on the minimum and maximum values of the variable. Negative skewness indicates the mean of a variable is less than the median, and the data is left-skewed. Positively skewed data is right-skewed with mean values larger than the median. The negative emotions have been divided into percentages of the total number of words in the review text. It can be seen that ratings were on average lower in 2020 than the year before. This difference is significant ($W = 6123227$, $p<0.000$) as concluded from a Mann-Whitney-U test.

4.1.1 Log transformations

The variables **nr words** and **nr emotional words** contain many outliers due to the differences in review length. The outlier values are far away from the mean. They have not been removed, but both variables were natural log transformed (see Appendix 13). This reduces the kurtosis and skewness in the data and results in a distribution of the variable that is as 'normal' as it can be. A feature *x* is replaced with log(*x*). A logarithm can be explained as log(*x*) = y because $x = b^y$. The base of the natural log is the mathematical constant "*e*" which is equal to 2.718282. The distribution of the feature remains the same. Continuous data is assumed to follow a bell curve when it is used for regression analysis. When this is not the case, the data can be log-transformed to de-emphasise outliers.

4.2 Descriptive analyses

To get a first insight into the main data, the descriptive statistics are investigated in more detail. It is checked whether the average polarity score aligned with the rating. In Figure 3, it is displayed that this was indeed the case. A rating of 1-star goes in line with a negative polarity score, and this does not significantly differ in 2019 and 2020. In Appendix 11 it is shown what reviewers are positive and negative about.

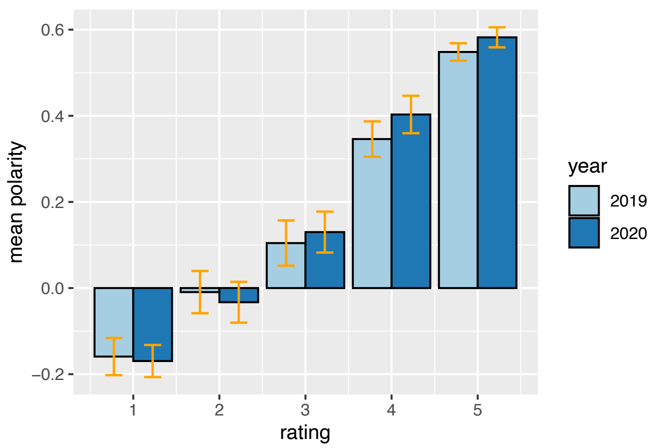


Figure 3. *Average polarity vs. rating score*

To understand seasonal trends, average polarity scores of 2019 and 2020 are compared per month. The average polarity score drops after March 2020, the month in which the effects of the coronavirus started to show in the United States (Tighe, 2021). Reviews were, on average, less positive in these months. This drop of average polarity in 2020 is not significantly different from 2019, as can be seen in the Figure 4 on the right, where the confidence intervals of the means are included.

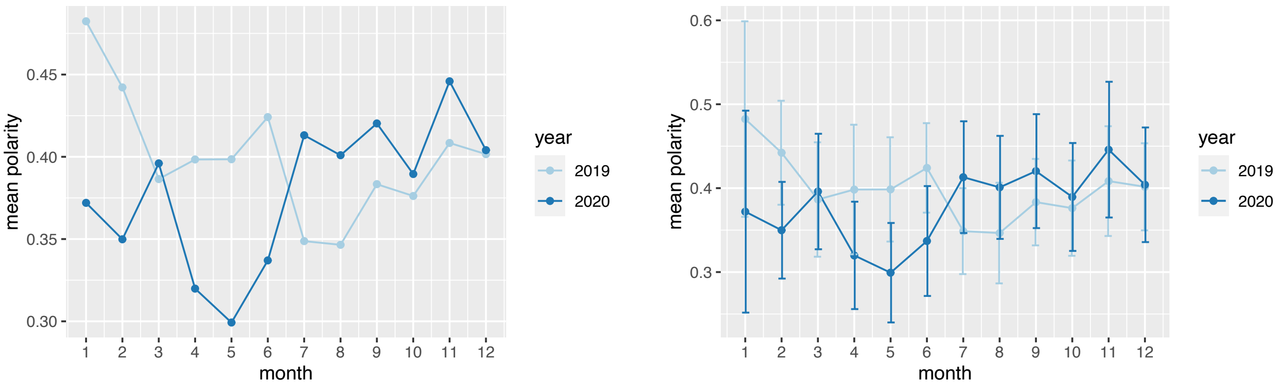


Figure 4. *Average polarity per month*

In 2020, the proportion of negative emotions had also been, on average, higher for the negative NRC emotions fear, sadness and anger. This difference in years is significant for sadness (W = 4864283, p<0.000) and anger (W= 4944924, p<0.001). Indicating fear, which is the treatment measure in the pre-test, is not significantly different from 2019 in 2020, but sadness and anger are.

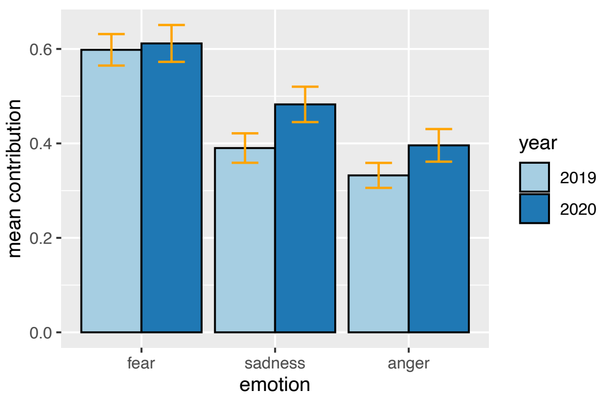


Figure 5. *Negative emotion division in 2019 and 2020*

Lastly, two descriptive analyses are performed to uncover differences in utilitarian and hedonic product types. It is checked whether the number of emotional wordings within the reviews and average ratings significantly differ in both product types. For all the reviews together, hedonic product received reviews contained on average more emotional wordings ($\bar{x}$ = 4.701, sd = 6.791) compared to utilitarian products ($\bar{x}$ = 3.408, sd = 3.961). The difference in rank sums is significant (W = 6446048, p<0.000). In Table 9, the differences between 2019 and 2020 are displayed. Hedonic products receive more emotionally loaded reviews than utilitarian in 2020.

Year	Type	Total words	Emotions	Percentage
2019	Hedonic	31,028	10,013	32.3
	Utilitarian	19,311	6,809	35.3

2020	Hedonic	25,456	8,608	33.8
	Utilitarian	19,680	6,025	30.6

Table 9. *Emotional wordings per year per product type*

Next, the average rating per product type is examined. Observations are summed up, and average differences are compared for both the data from 2019 and data from 2020. The average rating of utilitarian products is lower than those of the hedonic products, which is why the p-value is small even though the medians of the two groups are identical ($\tilde{x} = 5.000$). Ratings for hedonic and utilitarian product alternatives significantly differ in both 2019 (W = 1432356, p<0.000) and 2020 (W = 1166436, p<0.000). Due to the discrete data, sample means are not compared in these statistical tests. However, it is still checked how the means differ for both product types and years. The mean rating for utilitarian products is lower ($\bar{x}$=3.376) than for hedonic products in 2020 ($\bar{x} = 4.171$). The difference in average rank sums of the rating extremity is also significant in 2019 (W=1480949, p<0.01) and 2020 (W = 1098943, p<0.01), while hedonic products receive a more extreme rating in 2020. In 2019, this is the other way around. Utilitarian goods receive more extreme ratings.

4.3 Prediction models

To test the hypothesis of the conceptual model, four logistic regression models are developed, and one Random Forest model. They differ in use of the independent variables. Model 1 includes all features based on the sentiment in the review text. Model 2 contains only the review valence, polarity, and Model 3 all negative emotions. Lastly, Models 4 and 5 consist of all predictor variables and the interaction effect between emotional intensity and product type to test the moderating influence of product type on the relation between emotional intensity and product rating.

Model 1: *All sentiment*

$$Y_{RatingDummy} = \beta_0 + \beta_{LogNrEmotionalWords} + \beta_{Polarity} + \beta_{Fear} + \beta_{Sadness} + \beta_{Anger} + \beta_{EmotionalIntensity} + \varepsilon$$

Model 2: *Only polarity*

$$Y_{RatingDummy} = \beta_0 + \beta_{Polarity} + \varepsilon$$

Model 3: *Only negative emotions*

$$Y_{RatingDummy} = \beta_0 + \beta_{Fear} + \beta_{Sadness} + \beta_{Anger} + \varepsilon$$

Model 4 & 5: *All features*

$$Y_{RatingDummy} = \beta_0 + \beta_{LogNrWords} + \beta_{Fear} + \beta_{Sadness} + \beta_{Anger} + \beta_{Polarity} + \beta_{EmotionalIntensity} + \beta_{Type} + \beta_{EmotionalIntensity*type} + \beta_{Month} + \beta_{PriceLow} + \beta_{PriceMedium} + \varepsilon$$

Model 5 is the prediction model with a Random Forest and includes all variables and the interaction effect. With a model-based approach, it can very easily be derived which features are most important in predicting the rating. By indicating variable importance, it is tracked what changes are in model statistics for each predictor variable. The reduction in the statistic when the feature is added to the model is gathered. The total reduction is used as feature importance. In the Random Forest, for each tree, the prediction accuracy on the OOB portion is measured, and this is repeated after permuting each input feature. The difference between the two accuracies is averaged and normalised by the standard error, whereafter the most important feature can be detected.

4.4 Parameter tuning

To make the model predictions more accurate, hyperparameters should be tuned. These are parameters whose value is used to control the training process. If these are tuned, the optimal values are chosen, and the model can optimally solve the prediction problem. The loss function is minimised. There are no critical parameters to tune in logistic regression, but grid search can be performed with cross-validation. This is a resampling procedure to evaluate machine learning models. The single parameter *k* must be pre-determined to refer to the number of groups that the data sample is split into. This is done for the training data and set on 10, the default value. It controls for random effects as the train sample is cross-validated and optimised. This is also done for the Random Forest.

The other hyperparameters that result in the highest model accuracy for the Random Forest after predicting outcome variables with the train data are shown in Table 10. In Random Forests, only a random subset of features is considered per split. The range of subsets to be evaluated has been set to 1 till 10, and the random forest is trained over every possible value. For both years, 3 features are most optimal. Next, the maxnodes parameter is tuned. This indicates the number of terminal nodes in the forest. These are the nodes where the decision tree stops. The value at the terminal node is the value that the Random Forest predicted in that part of the tree. The optimal number is 20 nodes in 2019, and 21 in 2020. Lastly, the optimal number of trees are discovered.

	features	maxnodes	number of trees
2019	3	20	400
2020	3	21	200

Table 10. *Final Random Forest model hyperparameters*

The logistic regression models and the Random Forest model are compared. Model 4, the logistic regression model including all predictors, has the highest accuracy, Kappa statistic and precision scores in both 2019 and 2020. Therefore, it is concluded to be the best performing model in predicting the binary variable rating. Table 11 displays an overview of the model metrics.

Model 4 is used to interpret feature importance and draw conclusions on differences between features in 2019 and 2020. It can be concluded from the models that using review valence measured with polarity to perform sentiment analysis is a better-suited method compared to using separate negative emotions, or emotional intensity. Review valence (polarity) remains a significant and strong predictor over all models. This can be explained by the inclusion of positive valence in the polarity score, whereas negative emotions do not include positivity from the review and are worse in predicting positive ratings.

A possibility that Random Forest performed slightly worse than logistic regression can be overfitting. Decision trees are more likely to overfit the data since they split on multiple feature combinations. Logistic regression only associates with one parameter with each feature. Random Forest is an appropriate algorithm when there are sources of variation in the features, but in logistic regression feature selection was already applied. Due to the relatively imbalanced division of <5 and 5, the precision was consistently lower in 2019.

Logistic regression	Accuracy		Precision		Recall		Kappa statistic	
	2019	2020	2019	2020	2019	2020	2019	2020
Model 1: All emotions	0.654	0.652	0.503	0.558	0.660	0.750	0.291	0.322
Model 2: Only polarity	0.651	0.643	0.500	0.549	0.654	0.719	0.284	0.296
Model 3: Only NRC	0.687	0.650	0.576	0.585	0.391	0.500	0.256	0.261
Model 4: All	0.741	0.750	0.610	0.661	0.654	0.800	0.434	0.500
Random Forest								
Model 5: All	0.731	0.723	0.720	0.714	0.391	0.561	0.346	0.419

Table 11. *Model metrics*

5.1 Hypothesis testing

Fear of COVID-19 (H1)

H1, stating that COVID-19 has a direct negative impact on emotions is not supported by the small sample pre-test. Therefore, the hypothesis is rejected. The pre-test results showed no significant impact of the treatment condition on emotions, where respondents were reminded of their fear of the virus. However, it has been found that the treatment condition strengthens the impact of feeling very unhappy on the willingness to write a negative review. In literature it has been found that fear and sadness are direct consequences of a pandemic as COVID-19. In the main study, it became clear that in 2020, the proportions of negative emotions within reviews were on average higher. This difference is significant for sadness ($W = 4864283$, $p<0.000$) and anger ($W= 4944924$, $p<0.001$). This could be explained by people being more sensitive to emotions in 2020. Feelings of sadness increased and were expressed more, not necessarily fear.

Product type (H2a and H2b)

Hypothesis 2a, stating the number of emotional wordings is higher in times of a pandemic and higher for hedonic products compared to utilitarian products, is accepted. In 2019, on average $\bar{x} = 3.957$ emotionally loaded words have been used within the reviews ($sd = 5.470$), whereas in 2020 this was $\bar{x} = 4.224$ ($sd = 6.091$). The average rank sums are significantly different ($W = 4968621$, $p = 0.041$). Secondly, according to a Mann-Whitney-U test, the differences between the rank sums of utilitarian and hedonic use of emotional words were significant in 2019 ($W = 1767504$, $p<0.000$) and 2020 ($W = 1146620$, $p<0.000$). More emotional words have been used in reviews of hedonic products in 2019 ($\bar{x} = 4.803$, $sd = 6.712$) and 2020 ($\bar{x} = 4.704$, $sd = 7.156$) compared to utilitarian product reviews (2019: $\bar{x} = 3.095$, $sd = 3.611$; 2020: $\bar{x} = 3.667$, $sd = 4.492$). Therefore, it can be concluded that, on average, reviewers use more emotional words in a review written about a hedonic product alternative.

The effect of emotional intensity on product rating is not significantly strengthened by the **product type**. Therefore, hypothesis 2b is rejected. Product type is a categorical predictor. Out of the two levels, the hedonic product type is the referenced type, and the utilitarian type is inferred in comparison to the referenced variable. The interaction effect has been included in Models 4 and 5, which contained all features as predictors. There is no significant indication that product type is a moderator for the relationship between emotional intensity and product rating, and it does not strengthen the relationship in this study. Ratings were, however, more extreme for hedonic products in 2020, as concluded in the descriptive statistics ($W = 1211077$, $p<0.01$).

Negative emotions (fear, sadness, anger) (H3)

The third hypothesis of this study states that reviews with expressed negative emotions (fear, sadness and anger) are associated with a lower overall product rating and that this effect is stronger in times of a pandemic. Consumers who are very satisfied or dissatisfied are likely to post a review. In the pre-test, it had been discovered that respondents feeling unhappy during the pandemic are more likely to post a negative review (ρ= 0.10, p=0.05). Besides, all correlations between rating and negative emotions are significantly negative in both years (see Appendix 13). In 2019, fear had a slightly stronger negative correlation with rating compared to 2020, whereas in 2020 sadness had a stronger negative correlation with the rating. In Model 4 of the main study, only in 2020, **sadness** had a negative significant impact on the binary outcome variable rating, indicating that the more sadness is expressed in the review text, the less likely the rating is to be 5. From the coefficient of sadness ($\beta_{sadness}$ = -0.329, p <0.000) it can be concluded that when a reviewer expresses sadness with one more unit, the odds of the rating being 5 decreases with $e^{-0.329}$ = 0.719 -1 x 100% = -28.0%. The hypothesis is partially accepted.

Review valence and rating (H4, H5 and H6)

For hypothesis H4, it is examined whether reviews with a negative valence result in higher emotional intensity than positive valence. This is tested by taking a subset of the all the data based on the review valence, the **polarity** score. A negative polarity score results in a subset of 1,393 observations with a negative valence, whereas a positive polarity score results in a separate subset containing 3,770 observations accounting for positive valence. The average emotional intensity is calculated per subset. The average rank sum of emotional intensity within the negative valence subset is significantly larger (W = 2515920, p=0.01) than the positive subset (negative: $\bar{x}$ = 0.353, sd = 0.250; positive: $\bar{x} = 0.338$, sd = 0.273). Therefore, the hypothesis is accepted.

For hypothesis H5, it is examined whether the review valence mediates the relation between emotional intensity and rating. The variables used to test this are **polarity**, **emotional intensity** and the binary outcome variable **rating**. Mediation is hypothesised as a causal chain where the variable emotional intensity affects polarity, affecting the rating. First, the influence of emotional intensity on the binary variable rating is analysed and interpreted. This impact is significant in 2020 (2019: $\beta_{EmotionalIntensity}$ = 0.044, p=0.222; 2020: $\beta_{EmotionalIntensity}$ = 0.447, p=0.007). After that, the effect of emotional intensity on polarity is measured in a simple OLS model. This time, the predictor turned out to be positively significant in both years (2019: $\beta_{EmotionalIntensity}$ = 0.085, p=0.032; 2020: $\beta_{EmotionalIntensity}$ = 0.094, p=0.047). Lastly, both polarity and emotional intensity are included as predictors in a logistic regression model predicting the rating to confirm that the mediator affects the dependent variable while controlling for the independent variable emotional intensity. Polarity has a significant positive influence on the outcome variable in both years, and therefore, hypothesis H5 is partially accepted. The total effect of emotional intensity in 2020 on the rating is significantly explained by polarity, but not in 2019, as emotional intensity is not a significant predictor for the rating in that year. However, the other predictors are still significant in the mediation analysis, indicating a partial mediation exists and other features might also be of influence.

Emotional intensity not being a significant predictor for the binary variable rating in 2019 indicates a change in emotional intensity in 2019 does not significantly change the odds of the rating being 5, whereas this is the case in 2020. If emotional intensity increases in 2020, the odds of the rating being 5 go up significantly. This is also the case in Model 4 including all features. Review valence mediates this effect.

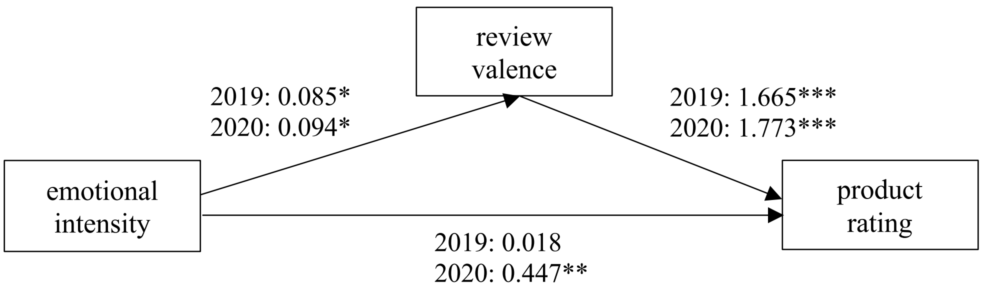


Figure 6. *Mediation of review valence*

Lastly, hypothesis H6 states review valence becomes more important in predicting the rating in times of a pandemic in 2020 compared to regular times in 2019. It can be seen in the output of Model 4, the model with the highest accuracy, that **polarity** has a significant positive influence on the binary outcome variable rating in both 2019 and 2020. The coefficients are respectively $\beta_{polarity}$ = 1.561; $\beta_{polarity}$ = 1.835. Polarity is a continuous feature; therefore, the log-odds are transformed to odds. When polarity increases with one, the odds that the rating is 5 increases with $e^{1.561}$ = 4.76 -1 x 100% = 376% in 2019, and $e^{1.835}$ = 6.26 -1 x 100% = 526% in 2020. It can be seen already that in 2020, the impact of polarity on the rating is larger than in 2019. This is, however, checked in more detail by uncovering the pseudo-R-squared. The R^2M is 0.161 in 2019 and 0.221 in 2020 for Model 4 with all features. A value between 0.2 – 0.4 indicates a very good model fit (McFadden et al., 1979). The importance of polarity can be discovered by comparing the average contribution of the feature on the total model predictions in a dominance analysis. In this analysis, one predictor is said to completely dominate another if its additional contribution to every possible model is greater than that of the other predictor (Azen & Traxel, 2009), by using the R^2M as fit function.

The average contribution of polarity has a value of 0.107 in 2019, and 0.141 in 2020. Hypothesis H6 is accepted.

Control variables

The two control variables added to the models with all features were **price category** and **month**. **Price low** and **price medium** had a significant positive influence on the product rating in both 2019 and 2020. This indicates that in 2019, when the price is in a low category, the odds of the rating being 5 are $e^{1.275}$ = 3.57 -1 x 100% = 257% larger than for the reference group **price high**, and for the medium category $e^{0.836}$ = 2.30 -1 x 100% = 130%. This is 276% and 201% for 2020, respectively. Both price dummy features are second and third most important in predicting the binary outcome variable rating in 2019 and 2020. **Month** had a significant negative influence on the dummy coded variable rating in 2019 (β_{month} = -0.048, p=0.001), and no significant effect in 2020. For 2019 this indicates that if the months increase with one, the odds of the product rating being 5 decreases with $e^{-0.048} - 1$ x 100% = - 4.68%. However, a drop was seen in the average polarity per month in 2020 after March. This drop was not significantly different from the average polarity score in 2019.

H1a	COVID-19 increases negative emotions experienced by consumers	Rejected
H2a	The number of emotional words in reviews is higher in times of a pandemic and higher for hedonic products compared to utilitarian products	Accepted
H2b	The product type strengthens the relationship between emotional intensity and rating	Rejected
H3	Reviews with expressed negative emotions (fear, sadness and anger) are associated with a lower overall product rating. This effect is stronger in times of a pandemic	Partially accepted
H4	Reviews with a negative valence contain a higher emotional intensity compared to neutral or positive valence	Accepted
H5	Review valence positively mediates the effect of emotional intensity of the review text on the product rating	Accepted in 2020
H6	Review valence is more important in predicting review ratings in times of a pandemic compared to regular times	Accepted

6. Conclusion and discussion

The main goal of this study has been to answer the research question of how emotional content in review text influences changes in the review ratings in times of a pandemic compared to regular times. The COVID-19 pandemic has had a large impact due to the worldwide lockdown and health consequences. Online buying behaviour increased. People were obliged to stay home for safety reasons, which caused a large number of customers to shift from offline to online shopping. Certain product categories became more popular, and previous research has concluded that consumers' emotions have been negatively affected by the virus. This study aimed to combine these two findings and determine whether this increase in negative emotions could also be measured among respondents in a small pre-test survey. It has also been examined how this translates into post-purchase evaluations. Reviews and ratings of U.S. reviewers have been analysed, sentiment analysis has been performed, and these sentiment scores have been used as predictors in multiple machine learning models.

The pre-test results showed that the treatment condition of fear of COVID-19 did not significantly directly affect emotions, but those feeling unhappy did show an increased willingness to write a negative review. The interaction between feeling unhappy and the treatment condition positively affects the willingness to write a negative review on a significant level, indicating the treatment condition strengthens this effect. Purchase intentions were higher for the utilitarian product alternative in the pre-test, but both utilitarian and hedonic product alternatives were included in the main study. The most prominent finding of the main study is that review valence and sadness became more important in predicting the rating in 2020 compared to the non-pandemic timeframe of 2019. Overall, more emotional words were used in 2020 and also, more emotional words appeared in reviews for hedonic products. Consumers tend to be more emotionally attached to these products. Besides, on average, more negative emotions were expressed in reviews in 2020 compared to 2019.

The product type did not significantly moderate the relationship between emotional intensity and the binary outcome variable rating. However, hedonic products did receive more extreme ratings. Four logistic regression models and one Random Forest model showed moderately high accuracies, precision, recall and Kappa statistics. These metrics were relatively most elevated for Model 4, the logistic regression model including all features. Therefore, this model has been used to draw conclusions on feature importance. Review valence, measured as polarity score, was the most important feature for predicting the binary outcome variable rating in both 2019 and 2020 and more important in 2020. This is checked by comparing the pseudo-R-squared which gives insight into the explainable power of logistic algorithms by comparing the (maximised) likelihood value from the current fitted model over the null model with only an intercept. Polarity measures the positivity or negativity of the review text. It moves in the same direction as the review rating. Polarity is also found to mediate the relationship between emotional intensity and review rating in 2020. It can be concluded that polarity explains the relationship between the independent variable emotional intensity and binary outcome variable rating in the year where the coronavirus was spreading. A partial mediation exists, while the coefficients of the features alone are also of significant impact in 2020. Emotional intensity was a more important and significant predictor for the binary variable rating only in 2020. This can be explained by the overall higher level of emotionality in this year, compared to the year before, where emotional intensity was not as important.

Besides determining sentiments with a polarity score, the NRC lexicon has been used to measure the negative emotions fear, sadness and anger. They have been used as predictor variables in the main study. In Model 1, the model with all emotions, they turned out to have a significant influence on the binary outcome variable rating, but this became non-significant in 2019 in Model 4, including all features. Interesting to see is that in 2020, sadness is the only significant negative emotion predictor. This can be derived from the descriptive statistics where it has been concluded that reviewers significantly experienced more sadness and anger in 2020. It can be concluded that consumers were not necessarily experiencing more fear, as expected in the pre-test, but were more sensitive to sadness and expressed this in their reviews. Sadness results in an increase in the amount people spent and should be investigated in more detail. Review valence being a more important predictor is likely to be caused by incorporating positive valence within the measurement of polarity. It turns out measuring negative emotions is not enough to cover changes in the review rating, compared to the review valence. However, a negative valence is not necessarily derived from negative emotions, while it can also be stressed from product weaknesses or problems. Therefore, it is important to consider both measurements.

A descriptive analysis of the division of polarity over 12 months in both years showed that after March 2020, the average polarity score dropped non-significantly for three months, whereafter, it rose again. Month had a significant negative influence on the product rating in 2019, which implies that an additional month decreases the odds of the rating being 5. The price categories low and medium have an important share in increasing the odds of the rating being 5 instead of <5 when compared to the reference variable of a high price category, because of the coefficient signs being positive in 2019 and 2020.

6.1 Practical Implications

The findings of this study result in opportunities for improvement for consumers, online rating communities, product owners and marketeers. By using the information on the drivers of emotionality in review text, negative or positive, marketeers can respond to this by communicating in an accumulative manner by increasing positivity and reducing negativity. The more positive the previous reviews, the more likely consumers are to engage in eWOM (Moe & Schweidel, 2012). Especially for the products used in this study, product owners should take the opportunity to shed more light on what reviewers are negative about in times of a pandemic. Sentiment analysis is a very suitable method to uncover such information by unnesting words out of the review text and labelling them or simply counting word frequencies. For product owners, competitive advantages can arise by comparing word sentiments of their product to the competitive product alternative. Due to changes in shopping behaviour and preferences, this information is more important than before. Managers often have a wait and see approach (Kim, 2020). The positive or negative words associated with the brand provide insights into what drives consumers to choose the brand. Loyal consumers are most valuable, mainly when they write positive product reviews. Other consumers will consider other reviewers' post-purchase evaluations for their pre-purchase evaluations, and positive reviews are more likely to be seen as informative and valuable. One negative product review can already be damaging as they are read more carefully due to curiosity.

Especially in times of a pandemic, product owners and brand managers need to be prompt. Reviews containing negative emotions should be an incentive for product owners to respond quickly to changes in rating and influence the reviewer with marketing communication techniques. A pandemic is a rapid and life-changing event with large impacts on shopping patterns. Sentiment analysis is a highly effective tool for a business to draw conclusions on brand perception and evaluate customer attitudes and emotions towards a product. Sadness is the most important emotion in 2020, decreasing the product rating, whereas review valence is most important in predicting the binary product rating. An algorithm can be built that tracks the review valence and words labelled as sad or fearful in review text using trained models such as developed in this study. By gathering large amounts of data, these models will become more accurate. When a negative review is detected, a chatbot can pop up to engage the customer with the brand and provide customer service. The emotion of fear causes readers to conform with the norm. The type of negativity in the review is important to be uncovered, while many negative reviews will make large groups of readers more negative. Emotion detectors in the review platform will provide interesting insights into whether a review was indeed informative. A negative valence or multiple sad words decrease the product rating, and this needs to be uncovered quickly.

Review valence can easily be detected from the review text with a polarity measurement within the review page. Positive reviews should automatically be placed on top of the review page instead of just considering the review rating as a determinator of the review's valence. These reviews are then more likely for consumers to be considered during their pre-purchase evaluation. Especially when the product owner sells a hedonic product alternative, they should be aware of the increasing number of emotional reviews in the study as they can expect more emotional wordings and extreme ratings. Consumers with an emotional connection with the brand have a higher lifetime value, but negativity should be detected.

Due to the lockdowns and no opportunity to shop offline in 2020, the online communities are expected to keep growing as multiple consumers engaged with the e-commerce industry. This expected increase is a great start for improvements and testing with large numbers of data.

6.2 Academic contributions

This study has been innovative due to the inclusion of analysis on COVID-19 and its direct and indirect impact in online sharing communities. Even though it cannot be validly concluded that this has been the cause for the change in importance of emotionality in 2020, multiple findings on emotionality and product types contribute to academia by broadening the knowledge of e-WOM in such unknown, fast-changing and insecure times as during a pandemic. Separate studies on the direct impacts of COVID-19 on emotions and emotions on product evaluations have been combined and assessed with a large data sample of reviews from deliberated product categories. Products considered in this study have been selected based on changes in purchase behaviour in their category due to the coronavirus. The time frames have been chosen selectively to increase internal validity and caused to have the ability to make feasible comparisons over the same months and include seasonal effects in the main study.

6.3 Limitations and suggestions for future research

Despite the contributions of this study, it also coped with limitations yielding for potential future research. The pre-test of this study has only been conducted among respondents located in the Netherlands, whereas the main study focuses on reviews in the United States. There are considerable differences in net optimism and purchase behaviour across countries due to variances in restrictions due to the pandemic. Besides, within the U.S., cultural and political differences can also occur as the country is relatively big. This impacts the external validity of the pre-test, for it is not highly generalisable over countries. Future research is recommended on location-specific effects of COVID-19 on emotions and review ratings where, as a result, differences can be discovered. However, the findings are still generalisable over comparable community platforms, especially the practical implications on sentiment algorithms within the platforms. The size of the pre-test survey has not been big enough to find significant impacts of COVID-19 on emotions. This could be caused by the mild treatment and measurement of fear of COVID-19 among respondents. Emotions and feelings of sadness could have been encouraged more in order for respondents to answer the survey items according to the way they actually feel or felt. In upcoming studies, the treatment conditions should be distinctive and more accurate to real-life situations.

Lastly, more advanced sentiment analysis algorithms can be used in future research such as deep recursive models for semantic compositionality by using Stanford's sentiment treebank. Their dictionary of sentiment labels includes 215,154 phrases and have been trained on a recursive neural tensor network resulting in accurate sentiment analysis (Socher et al., 2013). For the current research, feature importance would have been too complex to interpret with a neural network.

References

Addo, P. C., Jiaming, F., Kulbo, N. B., & Liangqiang, L. (2020). COVID-19: fear appeal favoring purchase behaviour towards personal protective equipment. *The Service Industries Journal*, 40(7-8), 471-490.

Ahorsu, D. K., Lin, C. Y., Imani, V., Saffari, M., Griffiths, M. D., & Pakpour, A. H. (2020). The fear of COVID-19 scale: development and initial validation. *International journal of mental health and addiction*, 1-9.

Ajzen, I. (1985). From intentions to actions: A theory of planned behaviour. In *Action control* (pp. 11-39). Springer, Berlin, Heidelberg.

Allon, G., & Bassamboo, A. (2011). Buying from the babbling retailer? The impact of availability information on customer behaviour. *Management Science*.

Amalia, P., Mihaela, D., & Ionuţ, P. (2012). From market orientation to the community orientation for an open public administration: a conceptual framework. *Procedia-Social and Behavioral Science*

Arndt, J. (1967) Role of product-related conversations in the diffusion of a new product. *Journal of Marketing Research* 4, 291–5.

Arnold, M. J., & Reynolds, K. E. (2003). Hedonic shopping motivations. *Journal of Retailing*, 79(2), 77– 95.

Azen, R., & Traxel, N. (2009). Using dominance analysis to determine predictor importance in logistic regression. *Journal of Educational and Behavioral Statistics*, *34*(3), 319-347.

Babin, B. J., Darden, W. R., & Griffin, M. (1994). Work and/or fun: measuring hedonic and utilitarian shopping value. *Journal of consumer research*, *20*(4), 644-656.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of general psychology*, 5(4), 323-370.

BBC. (2020, March 2). Coronavirus: Panic buying Australians clear supermarket shelves. Retrieved from https://www.bbc.com/news/av/world-australia-51702409

Boutsouki, C. (2019). Impulse behaviour in economic crisis: A data driven market segmentation. *International Journal of Retail & Distribution Management*, 47(9), 974–996.

Bradburn N.M., (1969). The structure of psychological well-being. Chicago: Aldine.

Brooks, S. K., Webster, R. K., Smith, L. E., Woodland, L., Wessely, S., Greenberg, N., & Rubin, G. J. (2020). The psychological impact of quarantine and how to reduce it: rapid review of the evidence. *The lancet*, *395*(10227), 912-920.

Burleson, Brant R. (2003), "The Experience and Effects of Emotional Support: What the Study of Cultural and Gender Differences Can Tell Us About Close Relationships, Emotion, and Interpersonal Communication," *Personal Relationships*, 10, 1, 1–23.

Chen, M. J. (2020). Examining the Influence of Emotional Expressions in Online Consumer Reviews on Perceived Helpfulness. *Information Processing & Management*, *57*(6), 102266.

DeVellis, R. F. (2003). Scale development: Theory and applications (2nd ed.). Thousand Oaks, CA: Sage.

Dodds, W. B., Monroe, K. B., & Grewal, D. (1991). Effects of price, brand, and store information on buyers' product evaluations. *Journal of marketing research*, *28*(3), 307-319.

Dhar, R., & Wertenbroch, K. (2000). Consumer choice between hedonic and utilitarian goods. *Journal of marketing research*, *37*(1), 60-71.

Fang, X., & Zhan, J. (2015). Sentiment analysis using product review data. *Journal of Big Data*, *2*(1), 1-14.

Fazio, R. H., & Petty, R. E. (2008). *Attitudes: Their structure, function, and consequences*. Psychology Press.

Fishbein, M., & Ajzen, I. (1977). Belief, attitude, intention, and behavior: An introduction to theory and research. *Philosophy and Rhetoric*, *10*(2).

Floh, A., Koller, M., & Zauner, A. (2013). Taking a deeper look at online reviews: The asymmetric effect of valence intensity on shopping behaviour. *Journal of Marketing Management*, *29*(5-6), 646-670.

Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information systems research*, *19*(3), 291-313.

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: the broaden-and-build theory of positive emotions. American psychologist, 56(3), 218.

Garg, N., & Lerner, J. S. (2013). Sadness and consumption. *Journal of Consumer Psychology*, 23(1), 106-113

Ge, F., Wan, M., Zheng, A., & Zhang, J. (2020). How to deal with the negative psychological impact of COVID-19 for people who pay attention to anxiety and depression. *Precision Clinical Medicine*, *3*(3), 161-168.

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. *International journal of endocrinology and metabolism*, *10*(2), 486.

Goes, P. B., Lin, M., & Au Yeung, C. M. (2014). "Popularity effect" in user-generated content: Evidence from online product reviews. *Information Systems Research*, *25*(2), 222-238.

Google Trends. 2021. *Google Trends*. [online] Available at: <https://www.google.com/trends/> [Accessed 20 May, 2021].

Grashuis, J., Skevas, T., & Segovia, M. S. (2020). Grocery shopping preferences during the COVID-19 pandemic. *Sustainability*, *12*(13), 5369.

Gupta, P., & Harris, J. (2010). How e-WOM recommendations influence product consideration and quality of choice: A motivation to process information perspective. *Journal of Business Research*, *63*(9-10), 1041-1049.

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet? *Journal of interactive marketing*, *18*(1), 38-52.

Hirschman, E. C., & Holbrook, M. B. (1982). Hedonic consumption: emerging concepts, methods and propositions. *Journal of marketing*, 46(3), 92-101.

Hovy, E. H. (2015). What are sentiment, affect, and emotion? Applying the methodology of Michael Zock to sentiment analysis. In *Language production, cognition, and the Lexicon* (pp. 13-24). Springer, Cham.

Holmes, E. A., O'Connor, R. C., Perry, V. H., Tracey, I., Wessely, S., Arseneault, L., et al. (2020). Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. *Lancet Psychiatry* 7, 547–560. doi: 10.1016/S2215-0366(20)30168- 1

Hu, N., Zhang, J., & Pavlou, P. A. (2009). Overcoming the J-shaped distribution of product reviews. *Communications of the ACM*, 52(10), 144.

Huang, A. H., Chen, K., Yen, D. C., & Tran, T. P. (2015). A study of factors that contribute to online review helpfulness. *Computers in Human Behaviour*, *48*, 17-27.

INSEAD. (2020). The Psychology Behind coronavirus panic buying. https://knowledge.insead.edu/node/ 13451/pdf

Jack, L., & Tsai, Y. D. (2015). Using text mining of amazon reviews to explore user-defined product highlights and issues. In *Proceedings of the International Conference on Data Science (ICDATA)* (p. 92). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).

Jiang, Z., & Benbasat, I. (2007). The effects of presentation formats and task complexity on online consumers' product understanding. *Mis Quarterly*, 475-500.

Johnson, E. J., & Tversky, A. (1983). Affect, generalization, and the perception of risk. *Journal of personality and social psychology*, *45*(1), 20.

J.P. Morgan, (2020, November 23). *How COVID-19 Has Transformed Consumer Spending Habits*. Retrieved from J.P.Morgan: https://www.jpmorgan.com/solutions/cib/research/covid-spending-habits

Keller, K. L. (1993). Conceptualizing, measuring, and managing customer-based brand equity. *Journal of marketing*, *57*(1), 1-22.

Kelly, H. (2011). The classical definition of a pandemic is not elusive. *Bulletin of the World Health Organization*, *89*, 540-541.

Kim, J., & Gupta, P. (2012). Emotional expressions in online user reviews: How they Influence consumers' product evaluations. *Journal of Business Research*, *65*(7), 985-992.

Kim, R. Y. (2020). The impact of COVID-19 on consumers: Preparing for digital sales. *IEEE Engineering Management Review*, *48*(3), 212-218.

Kronrod, A., & Danziger, S. (2013). “Wii Will Rock You!” The Use and Effect of Figurative Language in Consumer Reviews of Hedonic and Utilitarian Consumption. *Journal of Consumer Research*, 40(4), 726–739. doi:10.1086/671998

Kwartler, T. (2017). *Text mining in practice with R*. John Wiley & Sons

Laato, S., Islam, A. N., Farooq, A., & Dhir, A. (2020). Unusual purchasing behaviour during the early stages of the COVID-19 pandemic: The stimulus-organism-response approach. *Journal of Retailing and Consumer Services*, *57*, 102224.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical

Langhe, B. de, Fernandes, D., & van Osselaer, Stijn M.J. (2011). The anchor contraction effect in international marketing research. *Journal of Marketing Research*, 48(2), 366–380.

Lee, K. T., & Koo, D. M. (2012). Effects of attribute and valence of e-WOM on message adoption: Moderating roles of subjective knowledge and regulatory focus. *Computers in Human Behaviour*, *28*(5), 1974-1984.

Lerner, J. S., & Keltner, D. (2001). Fear, anger, and risk. *Journal of personality and social psychology*, *81*(1), 146.

Leverin, A., & Liljander, V. (2004). Journal of business & Industrial marketing. *Journal of Business & Industrial Marketing Management Decision International Journal of Bank Marketing Iss Journal of Product & Brand Management*, 19(11), 5–14.

Li, Y. (2015). Impact of impulsive buying behaviour on post impulsive buying satisfaction. *Social Behaviour and Personality: an international journal*, *43*(2), 339-351.

Lu, L. C., Chang, W. P., & Chang, H. H. (2014). Consumer attitudes toward blogger's sponsored recommendations and purchase intention: The effect of sponsorship type, product type, and brand awareness. *Computers in Human Behaviour*, *34*, 258-266.

Lu, J., Liu, Z., & Fang, Z. (2016). Hedonic products for you, utilitarian products for me. *Judgment & Decision Making*, *11*(4).

McFadden, D. (1973). Conditional logit analysis of qualitative choice behaviour.

McFadden, D. (1987). Regression-based specification tests for the multinomial logit model. *Journal of econometrics*, *34*(1-2), 63-82.

McFadden, D., Hensher, D. A., & Stopher, P. R. (1979). Behavioural travel modelling. *Croom Helm, London*, 279-318. Mehta, S., Saxena, T., & Purohit, N. (2020). The New Consumer Behaviour Paradigm amid COVID-19: Permanent or Transient?. *Journal of Health Management*, *22*(2), 291-301.

McKinsey & Company. (2020, October 26). *Consumer sentiment and behaviour continue to reflect the uncertainty of the COVID-19 crisis*. From Our insights: https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/a-global-view-of-how-consumer-behaviour-is-changing-amid-covid-19

Merenda, P. F. (1997). A guide to the proper use of factor analysis in the conduct and reporting of research: Pitfalls to avoid. *Measurement and Evaluation in counseling and Development*, *30*(3), 156-164.

Mitra, A. 1995. "Price Cue Utilization in Product Evaluations: The Moderating Role of Motivation and Attribute Information," *Journal of Business Research* (33), pp. 187-195.

Moe, W. W., & Schweidel, D. A. (2012). Online product opinions: Incidence, evaluation, and evolution. *Marketing Science*, *31*(3), 372-386.

Moore, S. G. (2012). Some Things Are Better Left Unsaid: How Word of Mouth Influences the Storyteller. *Journal of Consumer Research*, 38(6), 1140–1154. doi:10.1086/661891

Moore, S. G. (2015). Attitude predictability and helpfulness in online reviews: The role of explained actions and reactions. *Journal of Consumer Research*, *42*(1), 30-44.

Morgan, B. (2019). 100 stats on digital transformation and customer experience. *Forbes*.

Morning Consult (2020). National Tracking Poll# 200395. *Morning Consult: Washington, DC, USA*, 1-38.

Mudambi, S. M., & Schuff, D. (2010). What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com. *Management information systems*, 34(1), 185–200.

Pedrosa, A. L., Bitencourt, L., Fróes, A. C. F., Cazumbá, M. L. B., Campos, R. G. B., de Brito, S. B. C. S., & e Silva, A. C. S. (2020). Emotional, behavioural, and psychological Impact of the COVID-19 Pandemic. *Frontiers in psychology*, *11*.

Peng, C.-H., Yin, D., Wei, C.-P., & Zhang, H. (Eds.) 2014. How and When Review Length and Emotional Intensity Influence Review Helpfulness: Empirical Evidence from Epinions.com.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In Theories of emotion (pp. 3-33). Academic press.

Racherla, P., & Friske, W. (2012). Electronic Commerce Research and Applications Perceived 'usefulness' of online consumer reviews: An exploratory investigation across three services categories. *Electronic Commerce Research and Applications*, *11*(6), 548-559.

Ren, J., & Nickerson, J. V. (2019). Arousal, valence, and volume: how the influence of online review characteristics differs with respect to utilitarian and hedonic products. *European Journal of Information Systems*, *28*(3), 272-290.

Rozin, P., Berman, L., & Royzman, E. (2010). Biases in use of positive and negative words across twenty natural languages. *Cognition and Emotion*, *24*(3), 536-548.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and social psychology review,* 5(4), 296-320.

Russell, J. A. (1979). Affective space is bipolar. Journal of Personality and Social Psychology, 37(3), 345.

Russell, J. A. (1980). A circumplex model of affect. Journal of Personality and Social Psychology, 39(6), 1161.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. Psychological Review, 110(1), 145.

Shaver, P., Schwartz, J., Kirson, D., & O'connor, C. (1987). Emotion knowledge: further exploration of a prototype approach. *Journal of personality and social psychology*, *52*(6), 1061.

Sheth, J. (2020). Impact of Covid-19 on consumer behaviour: Will the old habits return or die?. *Journal of Business Research*, *117*, 280-283

Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C. (2013, October). Recursive deep models for semantic compositionality over a sentiment treebank. In Proceedings of the 2013 conference on empirical methods in natural language processing (pp. 1631-1642).

Tighe, D. (2021, March 9). *In general, has coronavirus impacted your shopping behaviour?* Retrieved from Statista.com: https://www.statista.com/statistics/1111366/coronavirus-impact-on-shopping-behaviour-us/

Valaskova, K., Kramarova, K., & Bartosova, V. (2015). Multi criteria models used in Slovak consumer market for business decision making. Procedia Economics and Finance, 26, 174-182.

Voss, K. E., Spangenberg, E. R., & Grohmann, B. (2003). Measuring the hedonic and utilitarian dimensions of consumer attitude. *Journal of marketing research*, *40*(3), 310-320.

Wang, J., Zhu, R., & Shiv, B. (2012). The lonely consumer: Loner or conformer? *Journal of Consumer Research,* 38(6), 1116-1128.

Yin, D., Bond, S. D., & Zhang, H. (2014). Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. MIS quarterly, 38(2), 539-560.

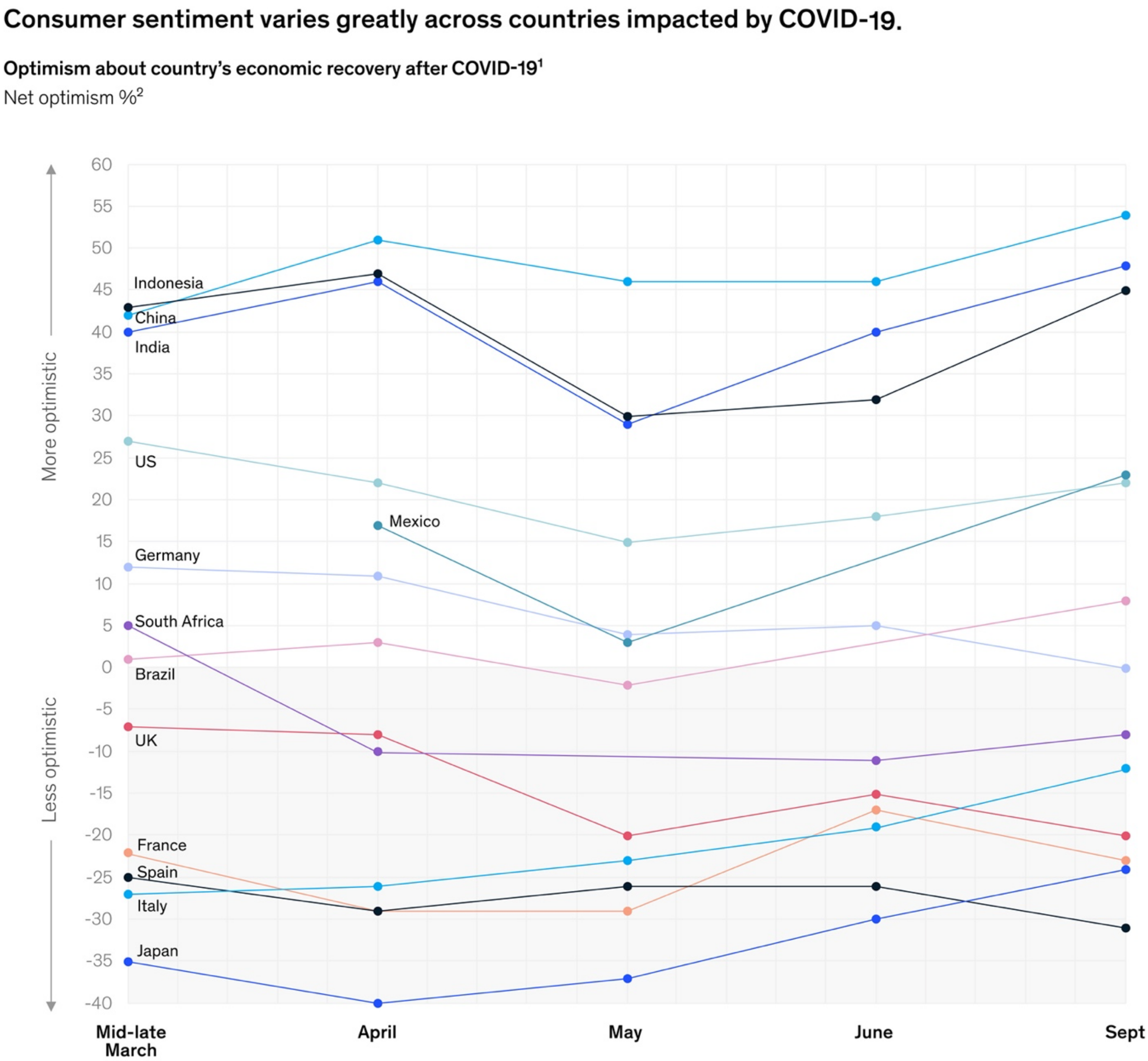
Yu, C., & Bastin, M. (2010). Hedonic shopping value and impulse buying behaviour in transitional economies: A symbiosis in the Mainland China marketplace. *Journal of Brand Management*, 18(2), 105– 114

Zablocki, A., Makri, K., & Houston, M. J. (2019). Emotions within online reviews and their influence on product attitudes in Austria, USA and Thailand. *Journal of Interactive Marketing*, *46*, 20-39.

Zhang, K. Z., Zhao, S. J., Cheung, C. M., & Lee, M. K. (2014). Examining the influence of online reviews on consumers' decision-making: A heuristic–systematic model. *Decision Support Systems*, *67*, 78-89.

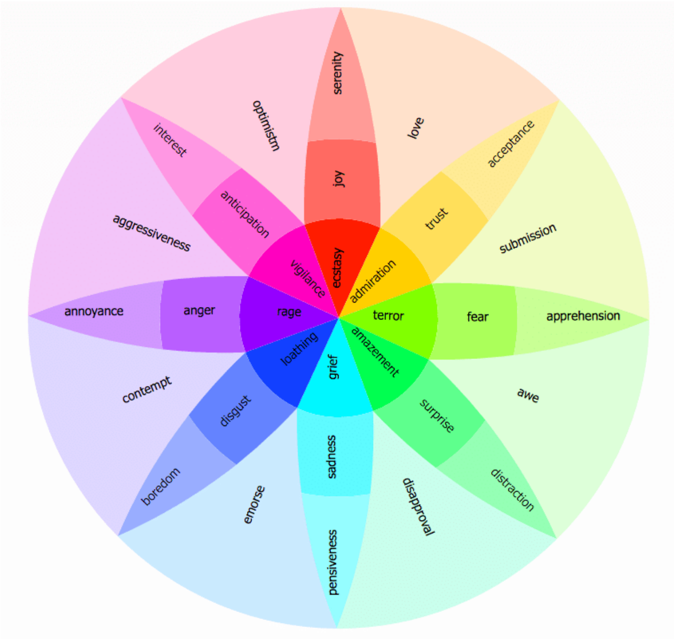
Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of marketing*, *74*(2), 133-148.

Appendix 1: COVID-19 worldwide sentiment



Source: McKinsey & Company COVID-19 Consumer Pulse Surveys, conducted globally between March 15-September 30, 2020

Appendix 2: Plutchick's wheel of emotions



Source: Plutchik, R (1980)

Appendix 3: Pre-test questions

<u>Part 1 Demographic information</u>

Q1) What is your gender?

Female

Male

Rather not say

Q2) What is your age?

< 18

18 – 34

35 – 64

Q3) How often do you do online shopping?

Never

About once every 6 months

About once every month

About once a week

More than once a week

Q4) Did your online shopping frequency increase due to the Covid-19 pandemic?

Strongly disagree

Disagree

Undecided

Agree

Strongly agree

Part 2: Fear of Covid – TREATMENT CONDITION

Participants answer: "Strongly Disagree", "Disagree", "Undecided", "Agree", "Strongly Agree" to the following:

Please go back to the start of the lockdown in 2020 and answer the following questions concerning your feelings about COVID-19

Q5_1) I was afraid of COVID-19.

Q5_2) It made me uncomfortable to think about COVID-19

Q5_3) My hands became clammy when I think about COVID-19

Q5_4) I have been afraid of losing my life because of COVID-19.

Q5_5) When watching news and stories about COVID-19 on social media, I became nervous or anxious.

Q5_6) I had trouble sleeping because I was worrying about getting the coronavirus.

Q5_7) My heart raced when I thought about getting the coronavirus.

Part 3: Current emotional states

Participants answer "Yes" or "No" to the following:

During the past months (did you feel)…

Q6_1) Particularly excited or interested in something?

Q6_2) Proud because someone complimented you on something you had done?

Q6_3) Pleased about having accomplished something?

Q6_4) On top of the world?

Q6_5) That things were going your way?

Q6_6) So restless that you couldn't sit long in a chair?

Q6_7) Very lonely or remote from other people?

Q6_8) Bored?

Q6_9) Very unhappy?

Q6_10) Upset because someone criticized you?

Part 4: Purchase intentions of hedonic and utilitarian products

Imagine that you have to work from home 5 days a week and spend most of your free time resting, watching TV. Lately you feel like you are tired of sitting on the same chairs. Would you buy a new desk chair or a new lounge chair?

Please answer the following questions about your purchase intentions

Note: it's not necessarily about what these examples look like

Desk chair

1. I would consider buying this product.
2. I have no intention to buy this product. (R)
3. It is possible that I would buy this product.
4. I will purchase this chair the next time I need one.
5. If I am in need, I would buy this chair.

Part 4a: Willingness to write review

Assume you bought this chair. When it arrives, you recognize its broken.

Please answer the following questions:

1. I am interested in sharing my negative experience with this product online
2. I am willing to write a negative product review online

After a few days using the product, you feel satisfied.

Please answer the following questions:

3. I am interested in sharing my negative experience with this product online
4. I am willing to write a negative product review online

Imagine that you have to work from home 5 days a week and spend most of your free time resting, watching TV. Lately you feel like you are tired of sitting on the same chairs. Would you buy a new desk chair or a new lounge chair?

Imagine that you have to work from home 5 days a week and spend most of your free time resting, watching TV. Lately you feel like you are tired of sitting on the same chairs. Would you buy a new desk chair or a new lounge chair?

Please answer the following questions about your purchase intentions

Lounge chair

1. I would consider buying this product.
2. I have no intention to buy this product.
3. It is possible that I would buy this product.
4. I will purchase this chair the next time I need a one.
5. If I am in need, I would buy this chair.

Part 4b: Willingness to write review

Assume you bought this chair. When it arrives, you recognize its broken.

Please answer the following questions:

1. I am interested in sharing my negative experience with this product online
2. I am willing to write a negative product review online

After a few days using the product, you feel satisfied.

Please answer the following questions:

3. I am interested in sharing my negative experience with this product online
4. I am willing to write a negative product review online

Appendix 4: Pre-test statistics

Appendix 4a: Cronbach's alphas

	Nr of items	Cronbach's alpha
Fear of COVID-19	7	0.84
Positive affect ABS scale	5	0.62
Negative affect ABS scale	3	0.64

Purchase intention	5	0.79
Willingness to write a negative review	2	0.91
Willingness to write a positive review	2	0.95

Appendix 4b: Descriptive statistics items

	Mean	Median	St. Dev	Min	Max
Q5_1	3.333	3	0.977	1	5
Q5_2	3.022	3	1.055	1	5
Q5_3	2.022	2	0.753	1	4
Q5_4	1.711	2	0.787	1	4
Q5_5	3.178	3	1.154	1	5
Q5_6	2.000	2	1.022	1	4
Q5_7	2.067	2	0.889	1	4
Q6_1	0.837	1	0.371	0	1
Q6_2	0.750	1	0.435	0	1
Q6_3	0.891	1	0.312	0	1
Q6_4	0.228	0	0.422	0	1
Q6_5	0.674	1	0.417	0	1
Q6_7	0.285	0	0.453	0	1
Q6_8	0.673	1	0.471	0	1
Q6_9	0.239	0	0.429	0	1
Q7_1	3.500	4	1.022	1	5
Q7_2	2.576	2	1.051	1	5
Q7_3	3.587	4	0.916	1	5
Q7_4	3.054	3	0.930	1	5
Q7_5	3.783	4	0.753	1	5
Q8	2.935	3	1.117	1	5
Q9	3.000	3	1.079	1	5
Q10	2.641	2	0.944	1	5
Q11	2.717	2	1.041	1	5
Q12_1	2.565	2	1.198	1	5
Q12_2	3.391	4	1.119	1	5
Q12_3	2.565	2	1.132	1	5
Q12_4	2.337	2	0.998	1	4
Q12_5	2.837	3	1.243	1	5
Q13	2.946	3	1.152	1	5
Q14	3.000	3	1.158	1	5
Q15	2.522	2	0.989	1	5
Q16	2.567	2	1.019	1	5

Appendix 4c. How often do you shop online?

	Frequency	Percentage	Cumulative percent
Never	0	0	0.0
Once every 6 months	13	14.1	14,1
Once a month	56	60.0	75.0
Once every week	19	20.7	95.7
More than once a week	4	4.3	100.0
Total	92	100	

Appendix 4d. Did your shopping frequency increase due to COVID-19?

	Frequency	Percentage	Cumulative percent
Strongly disagree	2	2.2	2.2
Disagree	12	13.0	15.2
Neither agree nor disagree	13	14.1	29.3
Agree	56	60.9	90.2
Strongly agree	9	9.8	100.0
Total	92	100.0	

Appendix 4e. Average ABS emotion score for both treatment conditions

	treatment	no_treatment
Average ABS score	1.533	1.489
Median	2.000	2.000
Min	-2.000	-3.000
Max	4.000	4.000

Appendix 5: Principal component analysis purchase intentions

Purhcase intention desk chair	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Standard deviation	1.658	0.959	0.761	0.669	0.550
Proportion of Variance	0.550	0.184	0.116	0.089	0.060
Cumulative Proportion	0.550	0.734	0.849	0.939	1.000

Purchase intention lounge chair	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Standard deviation	1.922	0.748	0.556	0.521	0.402
Proportion of Variance	0.739	0.112	0.062	0.054	0.032
Cumulative Proportion	0.739	0.851	0.913	0.967	1.000

Appendix 6: Pearson correlations purchase intentions and affect scores

			Positive affect				
	Purchase intention desk chair	Purchase intention lounge chair	Q6_1)	Q6_2)	Q6_3)	Q6_4)	Q6_5)
Purchase intention desk chair	1						
Purchase intention lounge chair	0.14	1					
Q6_1) Particularly excited or interested in something?	0.13	0.05	1				
Q6_2) Proud because someone complimented you on something you had done?	0.20*	0.02	0.36**	1			
Q6_3) Pleased about having accomplished something?	0.10	-0.06	0.32**	0.44***	1		
Q6_4) On top of the world?	0.07	-0.01	0.24*	0.31*	0.19	1	
Q6_5) That things were going your way?	0.13	-0.14	0.13	0.08	0.20*	0.32**	1

**** p < 0.001, ** p < 0.01, * p < 0.05*

			Negative affect		
	Purchase intention desk chair	Purchase intention lounge chair	Q6_7)	Q6_8)	Q6_9)
Purchase intention desk chair	1				
Purchase intention lounge chair	0.14	1			
Q6_7) Very lonely or remote from other people?	0.16	0.03	1		
Q6_8) Bored?	0.11	0.04	0.39***	1	
Q6_9) Very unhappy?	0.27**	0.25**	0.50***	0.23*	1

**** p < 0.001, ** p < 0.01, * p < 0.05*

Appendix 7: Principal component analysis willingness to write a review

Review negative desk chair	Comp.1	Comp.2
Standard deviation	1.339	0.455
Proportion of Variance	0.896	0.104
Cumulative Proportion	0.896	1.000

Review positive desk chair	Comp.1	Comp.2
Standard deviation	1.367	0.363
Proportion of Variance	0.934	0.066
Cumulative Proportion	0.934	1.000

Review negative lounge chair	Comp.1	Comp.2
Standard deviation	1.369	0.356
Proportion of Variance	0.937	0.063
Cumulative Proportion	0.937	1.000

Review positive lounge chair	Comp.1	Comp.2
Standard deviation	1.386	0.283
Proportion of Variance	0.960	0.040
Cumulative Proportion	0. 960	1.000

Appendix 8: Willingness to write a review Wilcoxon Signed Rank test per product type

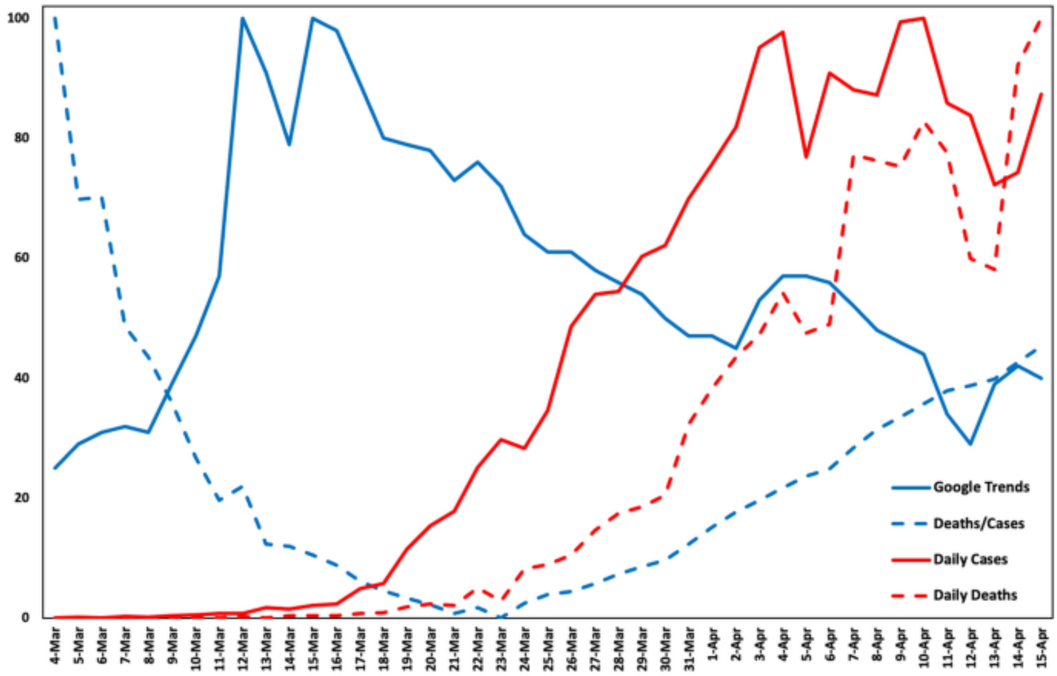
	V-statistic	p-value
Desk chair	769	0.004**
Lounge chair	980	0.000***

Appendix 9: Summary important statistics pre-test

Construct	Comparison	Test	Test-statistic	P-value
Shopping behaviour	Median of 2.5	Wilcoxon Signed Rank test	V = 4013	p = 0.000***
Fear of COVID-19	ABS – scale	Wilcoxon Rank Sum test*	W = 1089.5	p = 0.801
	Positive affect	*	W = 1167	p = 0.381
	Negative affect	*	W = 1082	p = 0.845
	Purchase intentions – desk chair	*	W = 1038.5	p = 0.884
	Purchase intentions – lounge chair	*	W = 1079/5	p = 0.866
	Willingness to write a negative review – desk chair	*	W = 1103	p = 0.713
	Willingness to write a positive review – desk chair	*	W = 1178	p = 0.328
	Willingness to write a negative review – lounge chair	*	W = 1174	p = 0.347
	Willingness to write a positive review – lounge chair	*	W = 1127	p = 0.571
Negative affect – item level (Q6_9 *Very unhappy?)*	Willingness to write a negative review	Pearson correlation	ρ = 0.10	p = 0.05
Purchase intention – desk chair	Willingness to write a negative review – desk chair	Pearson correlation*	ρ = 0.01	p = 0.910
	Willingness to write a positive review – desk chair	*	ρ = 0.11	p = 0.319
Purchase intention - lounge chair	Willingness to write a negative review – lounge chair	*	ρ = -0.13	p = 0.209
	Willingness to write a positive review – lounge chair	*	ρ = 0.04	p = 0.682
Willingness to write a negative review – desk chair	Willingness to write a negative review – lounge chair	Wilcoxon Signed Rank test*	V = 104	p = 0.676
Willingness to write a positive review – desk chair	Willingness to write a positive review – lounge chair	*	V = 227.5	p = 0.025*
Willingness to write a negative review – desk chair	Willingness to write a positive review – desk chair	*	V = 769	p = 0.004**
Willingness to write a negative review – lounge chair	Willingness to write a positive review – lounge chair	*	V = 980	p = 0.000***
Willingness to write a negative review – total	Willingness to write a positive review – total	Wilcoxon Rank Sum test	W = 3395	p = 0.017*

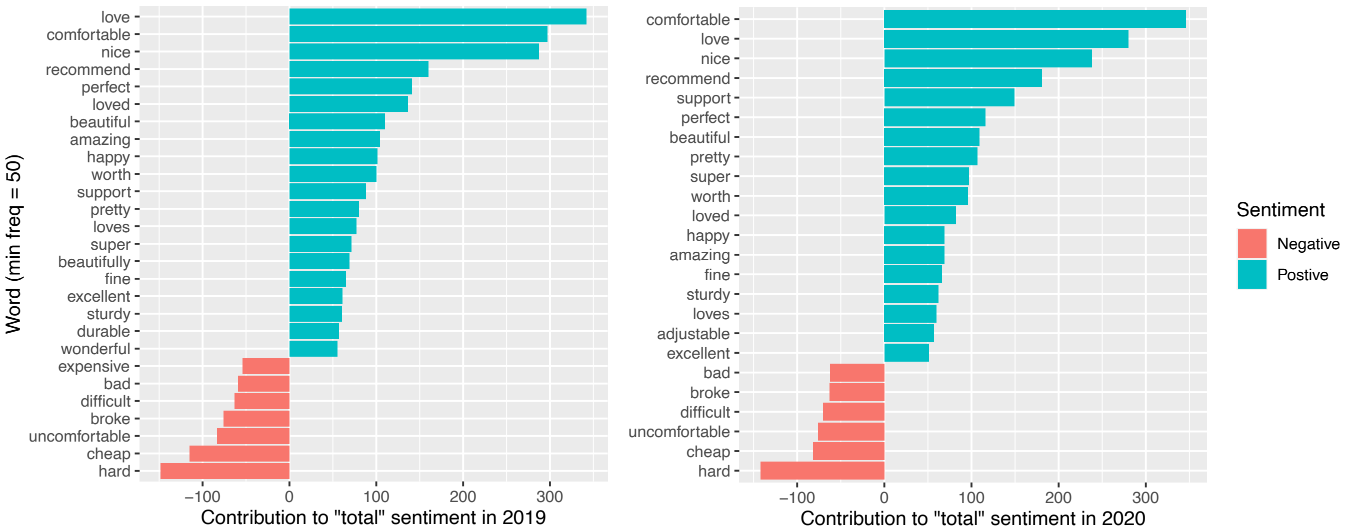
*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Appendix 10: COVID-19 search trend (Google trends, 2021)



Source: nature.com

Appendix 11: Positive and negative words used in 2019 and 2020



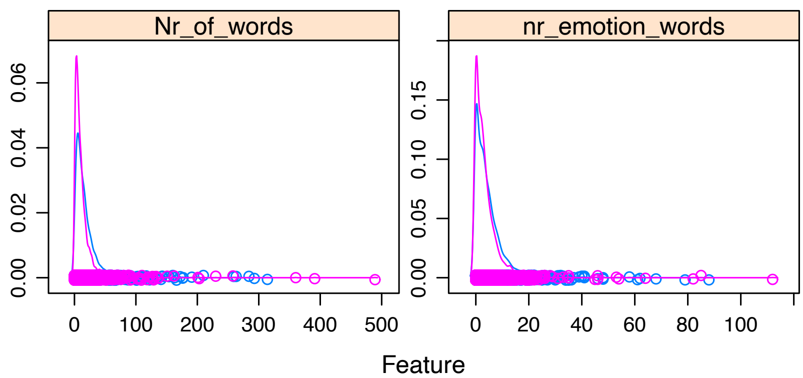
Appendix 12: Pearson correlations between features in main study 2019 / 2020

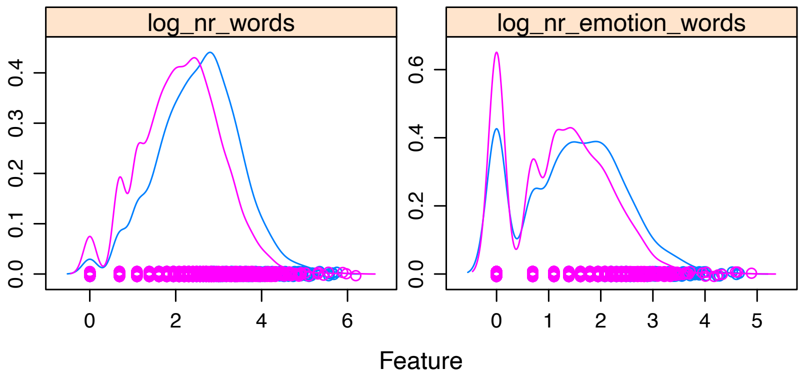
2019	1)	2)	3)	4)	5)	6)	7)	8)	9)
1) rating	1								
2) log nr words	-0.18***	1							
3)log nr emotion words	-0.12***	0.78	1						
4) polarity	0.43***	-0.06***	-0.05**	1					
5) fear	-0.13***	0.49***	0.62***	-0.20***	1				
6) sadness	-0.24***	0.50***	0.55***	-0.29***	0.69***	1			
7) anger	-0.18***	0.51***	0.57***	-0.23***	0.65***	0.68***	1		
8) rat.extr	0.42***	-0.15***	-0.09***	0.21***	-0.08***	-0.13***	-0.11***	1	
9) emotional_ intensity	0.01	0.09	0.58***	0.04*	0.30***	0.15***	0.17***	0.05**	1

2020	1)	2)	3)	4)	5)	6)	7)	8)	9)
1) rating	1								
2) log nr words	-0.21***	1							
3)log nr emotion words	-0.14***	0.77***	1						
4) polarity	-0.50***	-0.11***	-0.09***	1					
5) fear	-0.10***	0.47***	0.60***	-0.20***	1				
6) sadness	-0.27***	0.53***	0.58***	-0.29***	0.67***	1			
7) anger	-0.18***	0.50***	0.57***	-0.24***	0.70***	0.68***	1		
8) rat.extr	0.33***	-0.15***	-0.07***	0.19***	-0.07***	-0.13***	-0.09***	1	
9) emotional_ intensity	0.02	0.06**	0.56***	0.04*	0.28***	0.15***	0.17***	0.06*	1

**** p < 0.001, ** p < 0.01, * p < 0.05*

Appendix 13: Distribution after log transforming nr of words and nr emotion words





Appendix 14: Regression outputs

Model 1: All emotions

2019	Coefficients	Estimate	Std. error	z-value	p-value
	(Intercept)	0.485	0.084	5.809	0.000***
	log_nr_emotion_words	-0.350	0.084	-4.148	0.000***
	fear	0.190	0.072	2.744	0.008**
	sadness	-0.208	0.078	-2.644	0.008**
	anger	-0.019	0.087	-0.226	0.820
	polarity	1.529	0.097	15.769	0.000***
	emotional_intensity	0.7479	0.219	3.403	0.000***
2020					
	(Intercept)	0.186	0.092	1.981	0.04*
	log_nr_emotion_words	-0.428	0.091	-4.709	0.000***
	fear	0.364	0.078	4.664	0.001***
	sadness	-0.471	0.086	-4.925	0.000***
	anger	-0.052	0.086	-0.625	0.532
	polarity	1.761	0.108	16.059	0.000***
	emotional_intensity	0.857	0.252	5.081	0.000***

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Model 2: Only polarity

2019	Coefficients	Estimate	Std. error	z-value	p-value
	(Intercept)	0.315	0.046	6.786	0.000***
	polarity	1.568	0.089	17.450	0.000***

2020					
	(Intercept)	-0.07	0.052	-1.447	0.148
	polarity	1.856	0.099	18.026	0.000***

**** p < 0.001, ** p < 0.01, * p < 0.05*

Model 3: Only NRC (negative) emotions

Coefficients

2019		Estimate	Std. error	z-value	p-value
	(Intercept)	0.798	0.051	15.399	0.000***
	fear	0.147	0.065	2.317	0.01*
	sadness	-0.484	0.075	-6.400	0.000***
	anger	-0.203	0.081	-2.496	0.001***
2020					
	(Intercept)	0.572	0.056	10.133	0.000***
	fear	0.343	0.071	4.483	0.000***
	sadness	-0.776	0.079	-9.465	0.000***
	anger	-0.136	0.080	-1.665	0.09

**** p < 0.001, ** p < 0.01, * p < 0.05*

Model 4: All features

Coefficients

2019		Estimate	Std. error	z-value	p-value
	(Intercept)	0.541	0.231	2.285	0.022*
	log_nr_words	-0.414	0.065	-6.242	0.000***
	fear	0.173	0.080	2.155	0.031*
	sadness	-0.099	0.079	-1.465	0.215
	anger	-0.047	0.087	-1.814	0.595
	polarity	1.561	0.097	15.966	0.000***
	emotional_intensity	0.141	0.267	0.530	0.595
	type_utilitarian	0.460	0.153	30.001	0.002**
	emotional_intensity*type_utilitarian	-0.327	0.351	-0.932	0.351
	month	-0.048	0.015	-3.177	0.001*
	price_low	1.275	0.134	9.159	0.000***
	price_medium	0.836	0.243	6.210	0.000***

2020					
	(Intercept)	0.053	0.243	0.220	0.826
	log_nr_words	-0.408	0.071	-5.647	0.000***
	fear	0.194	0.089	2.178	0.02*
	sadness	-0.329	0.084	-3.710	0.000***
	anger	0.097	0.088	1.129	0.258
	polarity	1.835	0.108	16.746	0.000***
	emotional_intensity	0.573	0.235	1.957	0.050*
	type_utilitarian	0.101	0.165	0.609	0.542
	emotional_intensity*type_utilitarian	-0.861	-/397	-2.001	0.067
	month	0.01	0.016	0.994	0.322
	price_low	1.325	0.150	8.785	0.000***
	price_medium	1.104	0.142	7.746	0.000***

**** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$*

Appendix 15: Mediation analysis

1) Logistic regression, effect **emotional intensity** on binary outcome variable **rating**

	Coefficients				
2019		Estimate	Std. error	z-value	p-value
	(Intercept)	0.566	0.063	18.871	0.000***
	emotional_intensity	0.018	0.150	1.222	0.222
2020					
	(Intercept)	0.226	0.068	3.289	0.001**
	emotional_intensity	0.447	0.166	2.686	0.007**

2) Linear regression, effect **emotional intensity** on mediator **polarity**

	Coefficients				
2019		Estimate	Std. error	z-value	p-value
	(Intercept)	0.249	0.017	14.324	0.000***
	emotional_intensity	0.085	0.040	2.108	0.032*
2020					
	(Intercept)	0.266	0.020	12.828	0.000***
	emotional_intensity	0.094	0.049	2.001	0.047*

3) Logistic regression, mediating effect of **emotional intensity** and **polarity** on binary **rating**

	Coefficients				
2019		Estimate	Std. error	z-value	p-value
	(Intercept)	0.239	0.070	3.418	0.000***
	emotional_intensity	0.121	0.163	0.742	0.457
	polarity	1.665	0.095	17.452	0.000***
2020					
	(Intercept)	-0.171	0.070	-2.210	0.027*
	emotional_intensity	0.394	0.1844	2.136	0.032*
	polarity	1.773	0.09	17.872	0.000***

Appendix 16: Output dominance analysis with pseudo-R-squared 2019

2019: fitting 0 model for pseudo-r2

llh	llhNull	G2	McFadden	r2ML	r2CU
-1331.78	-1595.53	527.49	0.161	0.192	0.265

2019: Average contribution by predictor

	log_nr_ words	fear	sadness	anger	polarity	emotional intensity	type	month	price low	price medium	emotional intensity: type
r2.m	0.013	0.003	0.011	0.165	0.107	0.265	0.003	0.002	0.012	0.009	0.001

Appendix 17: Output dominance analysis with pseudo-R-squared 2020

2020: fitting 0 model for pseudo-r2

llh	llhNull	G2	McFadden	r2ML	r2CU
-1087.32	-1376.91	580.12	0.221	0.257	0.333

2019: Average contribution by predictor

	log_nr_ words	fear	sadness	anger	polarity	emotional intensity	type	month	price low	price medium	emotional intensity: type
r2.m	0.018	0.004	0.024	0.006	0.141	0.001	0.003	0.001	0.011	0.021	0.006

Appendix 18: Feature importance Random Forest