

# Exploring the Incongruence Between Product Descriptions and Customer Reviews to Complement Traditional Product Development

MASTERTHESIS

Isabel HERMES, 384949

February 26, 2020

ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

MSc Data Science and Marketing Analytics

Supervisor: Arash YAZDIHA

Second Assessor: Michel VAN DE VELDEN

*The views expressed in this thesis are those of the author and do not necessarily reflect those of the Erasmus School of Economics or Erasmus University Rotterdam.*

# Exploring the Incongruence Between Product Descriptions and Customer Reviews to Complement Traditional Product Development

Isabel HERMES, 384949

February 26, 2020

ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

MSc Data Science and Marketing Analytics

Supervisor: Arash YAZDIHA

Second Assessor: Michel VAN DE VELDEN

## Abstract

This research delivers an approach to examine the observed content differences between video game product descriptions and their respective customer reviews on Amazon. The aim is to better align the message sent by companies with what customers care about. Using human judges, it elicits prevalent features across product descriptions and matches these features based on whether they are mentioned in the respective customer reviews. MCA is used to establish the extent and structure of the mismatch. Thus, the existence of an incongruence between these text forms spanning the video game category is demonstrated. Given this, an individual product is singled out to explore the incongruence: *Call of Duty: Ghosts*. This sets an example of how a product manager might attempt to better align their product description with customer reviews. Aiming to use automation to resemble the task of human judges, LDA is used to generate topics in positively-rating reviews and contrasted with topics in negatively-rating reviews. The findings are used to recommend adjustments to the wording of the description and future product development decisions. The biggest issue customers complain about appears to be related to the servers. As suggested by the interpretation of LDA topics for the two corpora, the product description for the game in question should be adjusted to reflect well-received aspects and avoid problematic features. Product development should pay more attention to deliver reliable technical customer support and invest to improve the unsatisfactory parts of the game.

*Keywords:* customer reviews; product descriptions; product development; customer attitudes; means-end chain; multiple correspondence analysis; latent dirichlet allocation

*We've invented this wonderful system of language and calculation that is at once too simple to deal with the complexity of the world and also we are liable to confuse that system of symbols with the world itself, just as we confuse say money with wealth.*

ALAN WATTS

# CONTENTS

<b>1</b>	<b>Introduction</b>	<b>5</b>
<b>2</b>	<b>Literature Review</b>	<b>8</b>
1	The Role of the Customer in Product Development . . . . .	8
2	Theoretical Rationale . . . . .	11
2.1	The Role of Attitudes . . . . .	11
2.2	Consolidating Consumer Needs and Product Characteristics . . . . .	13
2.3	Informing Product Development . . . . .	14
<b>3</b>	<b>Establishing the Incongruence</b>	<b>16</b>
1	Data . . . . .	17
1.1	Selection of Product Group and Category . . . . .	17
1.2	Sampling Procedure . . . . .	18
1.3	Eliciting Product Attributes and Product Consequences from Descriptions . .	19
1.4	Matching Product Attributes and Product Consequences with Reviews . . .	22
1.5	Data Description . . . . .	24
2	Methodology - Multiple Correspondence Analysis . . . . .	25
3	Results . . . . .	27
3.1	Video Game Product Descriptions . . . . .	28
3.2	Video Game Customer Reviews . . . . .	31
3.3	Testing the Stability of the Data by Observing Bootstrapped Dimension Eigenvalues . . . . .	33
3.4	So what? Implications from the Existence of the Incongruence . . . . .	34
<b>4</b>	<b>Exploring the Incongruence</b>	<b>35</b>
1	Data . . . . .	36
1.1	Selection of Product to Examine . . . . .	36
1.2	Pre-Processing . . . . .	36
2	Methodology - Latent Dirichlet Allocation . . . . .	37
3	Results . . . . .	40
3.1	Choosing the Models . . . . .	40
3.2	Topics Inferred from Customer Reviews with Positive Ratings . . . . .	41
3.3	Topics Inferred from Customer Reviews with Negative Ratings . . . . .	45
3.4	LDA Topics in the Product Description . . . . .	49
3.5	So what? Implications for Description Formulation and Product Development	50
<b>5</b>	<b>General Discussion</b>	<b>54</b>
<b>6</b>	<b>Limitations and Further Research</b>	<b>56</b>

<b>7</b>	<b>References</b>	<b>60</b>
<b>A</b>	<b>Appendix: Amazon MTurk Survey</b>	<b>64</b>
<b>B</b>	<b>Appendix: Product Description of Call of Duty: Ghosts</b>	<b>65</b>
<b>C</b>	<b>Appendix: Illustrative sample of Input Data for MCA on 200 Video Game Descriptions</b>	<b>66</b>
<b>D</b>	<b>Appendix: Illustrative sample of Input Data for MCA on 1000 Video Game Reviews</b>	<b>67</b>

# 1. INTRODUCTION

The essence of selling goods besides ensuring quality is the communication of the product's worth to potential customers. For some goods, however, listing facts concerning the product's quality may not suffice. Such is the case for experience goods. It is a good whose value to a consumer only becomes apparent after its use (Nelson, 1970). Thus follows, that the message sent to potential customers must be convincing on several levels. Since the advent of the internet, it has become easier and therefore more common for consumers to provide feedback on their experience of using a product, often publicly. This is referred to as User Generated Content (UGC). Depending on the verdict, this feedback has the potential to be a blessing or a curse. Regardless, it is an important information resource for product development and marketing. On the one hand, favourable evaluations can confirm not only the product development team's ability to create a relevant product but also the marketing department's successful communication of that vision - ultimately bolstering sales. Conversely, unfavourable evaluations may lead to declining sales due to negative word-of-mouth (Dellarocas, 2003). The task then, is to utilize UGC to the company's advantage, effectively employing data mining to inform (future) product development and marketing operations. Indeed, it has been found that the top third of companies describing themselves as "data driven", exhibit on average 5% to 6% better performance than their peers (McAfee & Brynjolfsson, 2012). Exploring UGC and information about respective products as a combined source of content lends itself to several opportunities on the side of a producer. Based on the opinions of customers, valuable insights can be gained. These could be used to make marketing communications more relevant to the customer. Moreover, customers' observations can give relevant insights to product owners when deciding what to prioritize in development. It is thus of interest to investigate the content of UGC and product information in a combined effort; the aim is to explore whether this textual data from two different sources - consumers on the one hand versus producers on the other - can be leveraged to consolidate the product sold and the respective message sent by the producers. The research question is thus formulated as follows:

How can UGC be exploited to elevate the impact of marketing communications and inform product development?

The two-folded nature of this research question calls for the departments at hand to operate in line with one another. On the one hand, a product should satisfy the customer's needs and be of sufficient value to trigger repeat purchases. On the other hand, the message sent to the customer should communicate that that specific product is best suited for fulfilling those needs. A product is verbally interpreted twice. Firstly, by the copywriter, concocting the message sent to potential customers via various materials, including product descriptions. Secondly, by the consumer, expressing their opinion as UGC, through the use of social media or customer reviews. Successful communication is implied if the contents of these two agents' messages are in agreement. However, research suggests that a mismatch between the two can be observed (Ramsoy, 2015, p. 143). They observed that while the best product descriptions are widely believed to be

those listing factual information, customers respond more to positive or negative associations they might have with a product. The rationale from this is that the content of product descriptions is different from the content of customer reviews, implying a difference in message. By writing descriptions that appeal to positive associations customers have had with a product, potential customers might gain a more positive attitude towards purchasing the product. If the interplay between messages sent by producers and feedback given by customers is exploited effectively in the long term, this could lead to a more clear, favorable brand image. In the short term, next generations of products could be improved according to customer wishes and potential pitfalls could be avoided. This paper investigates this suggested incongruence by examining product descriptions and products reviews. By alleviating the incongruence between product descriptions and customer reviews for the company’s own product a lasting impact on the competitiveness could be achieved. Fundamentally, it is of interest to gather whether the associations voiced in customer reviews are in any way echoed in product descriptions. If mostly positive associations were present in descriptions, favourable views of a product could be enhanced. The presence of negative associations might deter customers from the product. This especially applies to products which are frequently re-published with improved functionality etc. Aiming to first establish whether this incongruence is actually observed in the real world, the first sub-question is formulated as follows:

*Sub-question 1:* Is there an incongruence between product descriptions and customer reviews?

The advantages of being able to analyze large amounts of data seem obvious. Due to the computational expense and the complexity of the data, it was common to conduct text analysis using human judges. However, using the computational power accessible today, the task of the human judge can be attempted to be automated using machine learning. Traditionally, approaches in this field are hybrids of human judgement combined with automation (Lee, 2007; Acher et al., 2012; Timoshenko & Hauser, 2019). It is necessary to conduct both a survey using human judge as well as employ machine learning to utilize the vast amounts of UGC-related data available. As such, this paper employs two distinguishable approaches: firstly, a human judgement approach and consequently a machine learning approach.

In order to tackle sub-question 1 - establishing the incongruence between product descriptions and customer reviews - a human judgement approach is used to analyze the content of product descriptions and customer reviews. It examines a number of different products of a product category, rather than a single product. This is crucial because the aim is to confirm the existence of a general incongruence for the product category at large, suggesting relevancy for all producing companies of this product. Employing humans in research is time intensive and costly. These constraints lead to a limitation of the number of texts to be examined, i.e. a small reviews-to-descriptions ratio. In the beginning, the human judgement approach elicits characteristics from product descriptions shared across the category at hand. Subsequently, the respective reviews of these products are judged as to whether they mention the elicited characteristics or not. Delving into the outcomes of these parts,

the prevalence and allocation of characteristics with respect to descriptions and reviews is examined to answer sub-question 1.

The goal of this paper is to establish whether UGC can be used to complement traditional product development through improving marketing communications. Thus, after establishing that there is an incongruence and therefore room for meaningfully exploring this incongruence, the machine learning approach tests the viability with regard to one product. Contrary to the human judgement approach, a large reviews-to-description ratio is employed. This is closest to what a product developer would be faced with in reality and demonstrates the efficiency of machine learning. The single product description is contrasted with the content of all its respective customer reviews. These findings are used to draw conclusions with respect to future products as well as current operations of the company e.g. customer service and the formulation of the product description at hand.

Accelerating the process of extracting information from text using the machine learning approach ideally assists both, the marketing as well as the product development department. It could exemplify what customers care most about regarding the product, shown by what they find relevant enough to include in their reviews. This can be used by marketeers to improve existing product descriptions. On the product development side, addressing specific feedback can steer future product development. The extent to which this is feasible in this setting will be addressed in sub-question 2:

*Sub-question 2:* How can customer reviews be used as grounds to adjust product descriptions and advise product development?

One of the goals in this research is relevancy to the business world. Therefore, product choice and research design have to be useful, repeatable and applicable to current products. The product category of video games is chosen. Firstly, video games are experiencing great popularity, and secondly, new versions are frequently released. There are a plethora of different games and therefore producers which have an interest in gaining advantages above their competitors, for example through heightened understanding of their customers. Furthermore, for video games, there is a large amount of information present on online shops, in form of product descriptions and customer reviews. Video games are experience goods, which means repeat purchases should be triggered, as is the case when looking at the publication of new versions of the same franchise every year. The contribution of this research is three-fold. First, it establishes a repeatable conceptual framework that can be used as the basis for different product categories. Second, it sheds light on the lack of coherence between product descriptions and customer reviews in the video game category, and therefore points toward a issue within that industry that might exist in other domains as well. Third, it develops a use case for examining and alleviating this incongruence and complementing product development for an individual video game, an experience good.



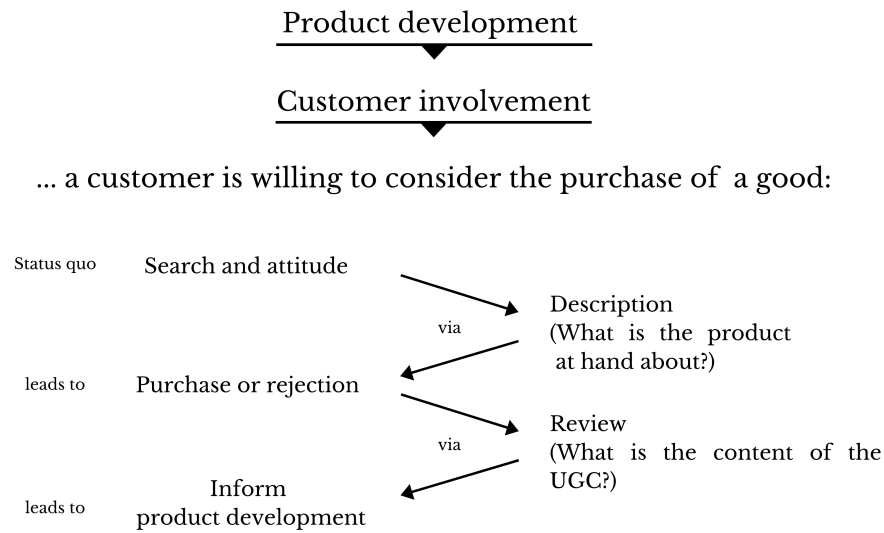
## 2. LITERATURE REVIEW

This section dives into relevant literature and establishes a theoretical background for the topic of this research: to enhance product development via bridging the incongruence between product descriptions and customer reviews. As such, rather than taking a holistic approach including all aspects of product development, the role of the customer is singled out. First, this relationship is explored. Then, the inner workings of how consumer attitudes, decisions and reactions influence the feedback loop from product sale, reiterating design aspects and development are laid out. This is reasoned using a theoretical framework. Lastly, to go full circle, this framework is related back to its function within the process of this research: enabling an exploration of the interplay between present marketing communication (via product descriptions) and the role of the customer (via customer reviews).

Before the role of the customer is discussed in more detail, the goal of selling the product in question has to be established. Mainly, the distinction between search and experience goods as established by Nelson (1970) must be made. Many video game production firms cyclically publish updated versions of their video game portfolio. This is steered by constant improvements in technological capabilities of consoles and high market demand for new gaming experiences. In order to stay competitive, game producers must keep up with the industry's developments. Thus, it is of interest to trigger as many repeat purchases as possible. Repeat purchases are purchases of the subsequent version of the preceding game, i.e. creating loyal customers. The goal of repeat purchases is a central characteristic of experience goods (Nelson, 1974); the act of purchase (or the rejection thereof) is the most direct path of influence from consumers to producers. The most distinct attribute distinguishing an experience good from a search good is, that the full value of an experience good can only be known after having purchased it. This holds for video games. However, it may be argued that video games are moving towards being search goods due to the prevalence of Youtube channels where players film themselves playing the game and display the game screen simultaneously. This arguably reveals big parts of what a video game is about. However, it decidedly withholds the actual experience and responsibility of playing the game by oneself. Therefore and in line with previous literature (Bragge & Storgårds, 2007; Schuff & Mudambi, 2012) this research treats video games as experience goods.

### 1. THE ROLE OF THE CUSTOMER IN PRODUCT DEVELOPMENT

Product development is informed by technological advances on the one hand and product performance on the market on the other. The role of the customer within product development may be indirect (implicit feedback by purchase decisions) or direct (involvement in development process through interviews). No matter the form of customer involvement, a product is irrelevant if it is delivered too late to the market and met with no demand. Thus, speed of development is of the essence: customer involvement should not negatively impact speed to market (Fang, 2008). Notwithstanding, customer input in the product development process may be of value. One stream



**Figure 2.1:** *Concept for overall understanding*

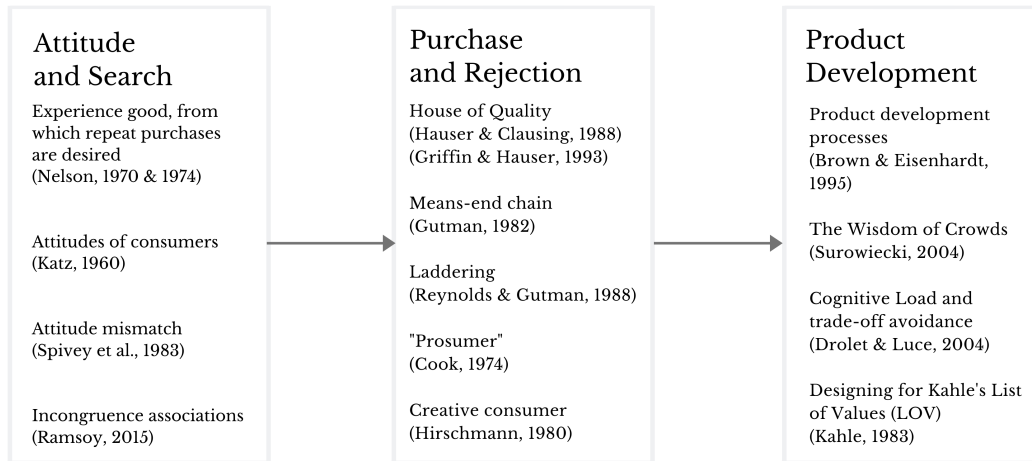
in product development is the rational plan (Brown & Eisenhardt, 1995). It postulates effective interplay of a relevant product plan, cross-functional communication and senior management approval as the prerequisite for a profitable product. Moreover, homogenous market information seems to inhibit novelty in products (Rindfleisch & Moorman, 2001). Indeed, it has been found that the more diverse the sources informing the product concept, the more innovative the end product (Amabile, 1983). Lettl (2007) investigates the role of customers in helping to create radical innovations. While the inclusion of so-called lead users in product development may lead to incremental innovation, radical innovation requires lead users with a technological understanding of the product to be innovated. Further, these customers must be motivated to use a new product with possibly radically different usability.

Nevertheless, incorporating customers in the development process can pose challenges. The rational plan includes early stage customer involvement in product development. Potential drawbacks of consulting individuals include "biases, memory lapses and myopia". These are made worse by a, "singularity of opinion" related to the small number of individuals consulted (Brown & Eisenhardt, 1995) due to cost and time constraints. Furthermore, customers could refuse participation and laying bare their thoughts. Thus, the main complications related to the role of the customer in product development are: hampering of speed to market and the lack of a reliable, uniform and consistently available opinion. To circumvent an explosion of costs, efficiency is of the essence. Amends can therefore be made by accelerating interrogation and evaluation and consulting a higher number of individuals. An effective undertaking of a customer-centric approach to product development, then, has the aforementioned prerequisites. The requirement of consulting a high number of individuals can be fulfilled by tapping into two resources available today: first, the large amounts of data and

second, advanced efficient computational tools. Using different kinds of UGC as data sources, Latent Dirichlet Allocation (LDA) has successfully been implemented to gauge aspects of the texts at hand (Brody & Elhadad, 2010; Ramage, Dumais & Liebling, 2010). Lee (2007) uses review data to elevate existent research on product attributes mentioned in UGC to construct a relation from attributes to user needs. Acher et al. (2012) used product descriptions to create Feature Models of products, that is, structured and broken down visualizations of core product functionalities and their relation to one another. This research investigates the interplay between product descriptions and customer reviews to assess what matters to customers through finding attributes in both texts and comparing their presence.

UGC in the form of customer reviews are an appropriate source of customer views (Lee, 2007; Brody & Elhadad, 2010; Lee & Bradlow, 2011; Timoshenko & Hauser, 2019). It has been shown that the mention of brand names in forum entries can be used to infer the market structure of an industry (Netzer, Feldman, Goldenberg & Fresko, 2012), based on associations and occurrences. Thus, looking at UGC has the potential to reveal not only what customers think but also a deeper, web-like structure. Timoshenko and Hauser (2019) demonstrated that in the product category of oral care, the number of needs established from UGC (customer reviews) was at least as high as the number of needs found in experiential customer interviews. Customer reviews can thus be regarded as sufficiently similar to the output of traditional customer interviews. Furthermore, the use of a high number of individual opinions for innovation is in line with the theory presented in, "The Wisdom of the Crowds" (Surowiecki, 2005). Using UGC enables an analysis of the market's (customers') expectations and attitudes towards an existing product. Customers' views and understanding of a product offering are influenced by marketing communication. Consolidating marketing communication with customer expectations of this product can manage expectations and be a useful tool to impact purchase decisions and ultimately improve feedback. One example of marketing communication are product descriptions containing information on product characteristics.

Coming from a product development approach, and singling out customer involvement as one aspect to utilize more effectively, the ground assumption is that a potential customer is willing to consider the purchase of a good. Searching for a suitable good to purchase, marketing communication regarding the product at hand influences their decision of purchase or rejection. In the case of a purchase, the experience with the product (potentially) leads to a customer expressing their opinion on the product, in form of a review. The content of UGC can then be utilized to help shape the development vision of the product created next. With respect to video game development, the findings of this research are most applicable to the pre-production (formulating product descriptions) and post-production (by looking at customer reviews) phases (Chandler, 2013). Thus, utilizing market information ideally enables valuable inputs informing both future product development as well as needed adjustments to present product information. An overview of the role of the customer and the approach taken in this research is seen in Figure 2.1.



**Figure 2.2:** *Theoretical rationale as per concept*

## 2. THEORETICAL RATIONALE

### 2.1 The Role of Attitudes

In order to use the role of the customer in product development, one must understand the thought process of a consumer once they are exposed to a product. The way they think about a product is important because that impacts whether they are willing to consider a purchase. In effect, we want to examine how the consumer forms a judgement, or opinion about a product. This is in line with Katz' (1960) assessment that opinions originate in attitudes, "Attitude is the predisposition of the individual to evaluate some symbol or object or aspect of this world in a favourable or unfavourable manner. Opinion is the verbal expression of an attitude, but attitudes can also be expressed in nonverbal behavior." In the scope of this paper, this statement helps justify the choice of text to represent attitudes, since verbal expressions are essentially opinions. Katz employs the example of how the voting public's allegiances to different political parties may be formed. This research extrapolates his study of attitudes and applies it to the process of a marketer (corresponding to a politician in Katz' example) selling a product (party) to a potential customer (voter). The power related to exploring attitudes to improve consumer behavior is apparent. Katz establishes four categories of attitudes: adjustment, value-expression, knowledge and ego-defense. This research focuses on the first two. The adjustment category is linked to utility and can thus be considered most goal-oriented. Individuals signaling this attitude display willingness to assess their surroundings based on how much closer something brings them to fulfill the need linked to their most valued goal. The expression of values on the other hand, concerns the individual's need for adhering to their established sense of self. The study of attitude arousal and attitude change involves complex operationalization (Locander & Spivey, 1978) and therefore, this research is limited to the theory behind. Whatever the attitude may be, consumer behavior depends most on the extent to which the stimulus at hand appears to further the individual's journey in a positive direction.

It is postulated that utilitarian (adjustment) attitudes are more robust to change than value-expressive attitudes. Intuitively, this can be explained by the nature of the relationship between product attributes and consequences (Spivey, Munson & Locander, 1983). While attributes are observable as a directly experienced characteristic fixture of the product in question, the consequences of product use experienced, may vary on the individual level. Attributes and consequences are discussed in more detail in Section 2.2 below. To take a closer look at how attitudes and emotions may influence consumers and their decisions, trade-off avoidance is an essential concept. Trade-off avoidance is triggered by an individuals' resistance to sacrificing one outcome representing a personal goal over another representing a conflicting personal goal. Irrationally so, the emotional unease experienced by having to make such choices and incurring the loss of the ignored option is mitigated by the use of bias and heuristics - rather than rational weighting of the options against one another.

There are conflicting findings regarding cognitive load and the human response to different extents of load. Cognitive load is the amount of sensory impressions an individual finds themselves exposed to (Sweller, 1988). Intuitively, the larger the load, the less attention they are able to allocate to processing each impression, thus becoming more susceptible to reacting randomly rather than rationally. It follows, that exposure to large cognitive loads implicates the ability to engage in rational assessment of different outcomes. Hence, trading off one outcome over the other is based more on one's momentary emotion and less on a complete view of behavior likely to lead to one's most preferred outcome. Contrary to this intuition, one study suggests that higher cognitive load leads to more normative decision making (Drolet & Luce, 2004). It finds that an increase in cognitive load can lead to a decrease in emotional trade-off avoidance because less cognitive resources can be allotted to emotions. This lessens the mental image of the potential "blow" to personal goals due to the choice against a certain opportunity.

Therefore, taking into account the possible effect of cognitive load on evaluations of a product could give insights regarding how product descriptions should be phrased. This can be in terms of whether the text should address rather utilitarian (attributes) or value-expressive (consequences) goals, but also in terms of the sheer amount of text, depending on what reaction is desired. That is, in the case of larger, more confusing texts, trade-off avoidance might be mitigated, while smaller texts might challenge the customer in terms of decision making and lead to less normative choices, since their cognitive ability is not impaired. In combination with cognitive load present in product descriptions, the aforementioned study by Ramsøy (2015) can be consulted to speculate possible improvements of said descriptions. It found that positive and negative associations had stronger effects on consumers than mere product facts. Assuming that product descriptions consist of factual listings of features and attributes, there is room for entering a more emotive rapport with consumers reading the product descriptions. Ideally, this would engage the consumer in ways that are more likely to lead to a purchase. Thus, it would be interesting to see whether this finding can be extrapolated with respect to associations found in customer reviews, with the goal being to exploit the content of

customer reviews to inform product development and product description marketing.

## **2.2 Consolidating Consumer Needs and Product Characteristics**

A way of incorporating customer views into product development is called the House of Quality (Hauser & Clausing, 1988). It uses customer needs and engineering standards to streamline the product development process, from idea generation to evaluating products on the market. The first house focuses on obtaining and evaluating customer needs (Griffin & Hauser, 1993). Identifying existing needs can help infer how well a product satisfies these needs and what the extent and characteristics of gaps in the current market are. In line with traditional marketing practice, Griffin and Hauser employ user interviews to find needs. During this lengthy and expensive process, users then illustrate their experiences with a product. The authors already comment on the issue of time delay and high expenses caused by conducting customer interviews in this style.

To advance product development beyond the traditional use of customer input, this paper suggests exploring User Generated Content (UGC) as a source of the Voice of Customer. Thus, in place of interviews, customer reviews are used and treated as transliterations. Voice of Customer was established as a concept by Griffin and Hauser (1993) in order to involve customers in the product development process. They define it as a, "hierarchical set of customer needs". In their research this hierarchy is fundamental to building the new product. On this note, it is important to note that it is common to distinguish between product attributes and customer needs (Lee, 2007). However, some literature suggests using the terms interchangeably (Krishnan & Ulrich, 2001). An alternative practice constructs a link between attributes and needs, or rather consumer values (Vriens & Ter Hofstede, 2000). They do this by employing the means-end chain in combination with Kahle's List of Values (LOV).

The means-end chain was developed by Gutman in 1982. The rationale behind this theory is that consumers have certain individual values that they ultimately strive to achieve. Companies aiming to sell their product to a select group of consumers must therefore understand how to signal to consumers that their product will help them reach their personal values. Thus, the product characteristics at the one end must match or lead to a person's individual (set of) values at the other end. In the end, a consumer aims to achieve their values. The product and its marketing have to signal to the consumer that with its purchase, achievement of their values can be approached. At the ground level, Gutman maps this connection by eliciting attributes from products. The attributes then lead to certain consequences: benefits (desired consequences). These are the means to reach the value: from a product consequence, the consumer is enabled to approach a certain value. This value is the end, thus the means-end chain connects the means (product with attributes) to the end (value) via consequences. One approach to elicit the means-end chain is by laddering (Reynolds & Gutman, 1988). It aims to uncover, define and connect the three levels of the means-end chain: attributes, consequences and values.

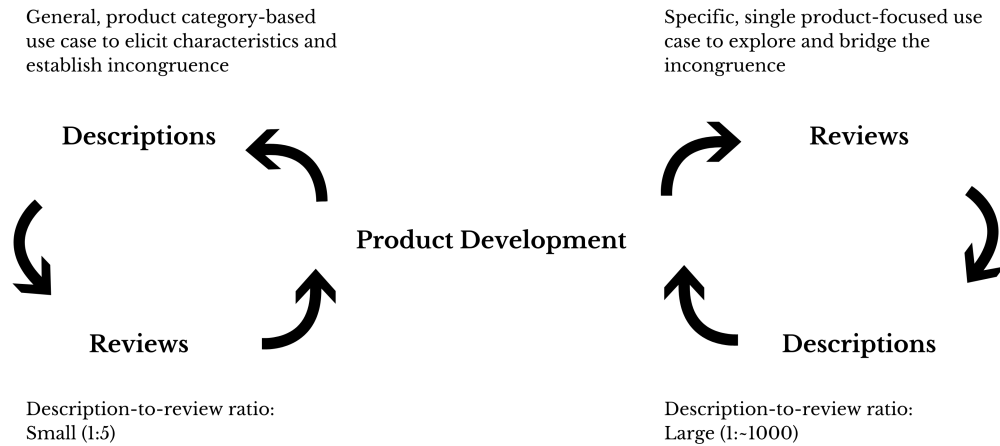
As with the House of Quality, laddering traditionally employs interviews. During these, interviewees are asked questions regarding a product, such as why they chose it. The answers to the initial questions often feature attributes. In order to elicit consequences of a certain product, the interviewer probes further how what the consumer mentioned fulfills certain capabilities and their motivations for purchase. Ultimately, all mentions of attributes, consequences and even values are highlighted in the transcribed text and put into hierarchical order, i.e. the ladder is constructed. To determine which attributes lead to which consequences and ultimately values, a repertory grid can be used. In their article, Vriens and Ter Hofstede (2000) lead the construction of the means-end chain by conducting interviews in this way asking users about their experience with a product. During each interview the interviewer starts with asking what attributes the consumer sees in the product and then continues by eliciting the benefits or consequences these attributes result in to ultimately probe the consumer as to what final value the product brings them based on these attributes and consequences. They suggest to assess values based on the 9 values included in Kahle's List of Values (LOV). They state, "(values are) universal to a very large extent". Thus, Kahle's LOV are used across consumers and products. These values are as follows: *being well respected*, *excitement*, *fun and enjoyment in life*, *security*, *self – fulfillment*, *self – respect*, *sense of accomplishment*, *sense of belonging* and *warm relationship with others*. In line with Kahle and Kennedy (1988) the values *excitement* and *fun and enjoyment in life* are gathered up in one category: *fun and enjoyment in life*. This is due to the fact that the former was seldom selected by Americans in a survey to be their most important value and if it was selected as such, the latter often followed as second most important value. Thus, there will be 8 values used in this research.

Based on the abovementioned theories, the first hypothesis is established relating to the left part of Figure 2.3:

H1: The product attributes and product consequences observed in video games display a different pattern in product descriptions than in customer reviews.

### 2.3 Informing Product Development

This research aims to employ UGC in order to elevate products brought to market. As such, Voice of Customer is used to inform product development and bridge the gap between existing products and what consumers desire. A term to describe the role of the customer in this is prosumer, coined by Alvin Toffler in his 1980 book "The Third Wave". By prosumer, he refers to those who not only consume but also produce their own goods. In a nutshell, the rationale is that on a spectrum where consumer and producer are imagined on the far ends, the prosumer is gradually moving away from being just a consumer and towards also being a producer - Toffler applied his theory on the Industrial Age (and what follows). While in the case of video games, customers do not literally



**Figure 2.3:** *Going Full Circle*

produce their own games, their opinions in form of UGC can have direct impact on the future game versions of the franchise. This especially holds when they voice an idea that is novel, creative and realizable. The concept of utilizing consumers' creativity was also examined by Hirschman (1980). She explored innovation, creativity and search for novelty in relation to consumers. In order for a consumer to be creative, she found there are two prerequisites relating to cognition. Firstly, the presence of life experience in relation to having consumed a number of products and being aware of their possible attribute combinations and secondly the ability to form a mental network to connect this experience into a pattern. Another study found that highly creative people are able to find solutions to challenges, i.e. are analytical and cogent (Welsh, 1975). Going full circle, consumers are also more likely to adopt innovative products (Hirschman, 1980). Referring back to the concept of Wisdom of the Crowds, the existence of a plethora of consumer views and ideas in UGC has the potential to inform traditional product development beneficially.

The combined mass of customer insights on a product (category) can be used to paint a picture representing how customers see the product (category). Whether this picture is coherent with the way it is marketed, is of interest because it might reveal shortcomings of the message sent to consumers. One part of this message is the product description. It is likely that neither all customer reviews are negative nor all are positive. Explicit ratings in the form of stars that can be rewarded out of 5 are common on e-commerce websites. Exploring whether views expressed in reviews that rated the product positively reveal different content than those that rated the product negatively is a second layer to this inspection. The hope is to extract knowledge about personal views and desires from customer reviews and to then employ that knowledge, reinforcing the relevance of the message sent to consumers by reformulating the product description in accordance. Based on the theories relating to the prosumer the following hypothesis is formulated in correspondence to the



right hand side section of Figure 2.3.

H2.1: Differences in content between positively and negatively-rating customer reviews are clear enough to inform a rephrasing of product descriptions.

Despite increasing the appeal and relevance of product descriptions, review content can be used for product development. When expressed in the form of e.g. wishes or complaints, consequences or weaknesses of the product can be identified. This can then be communicated to the product development and design teams and thus provide actionable insights for them to build upon. Therefore, such insights can be used to inform new product development or even innovation. The following hypothesis is related to the right hand side section of Figure 2.3:

H2.2: There is underlying information in customer reviews that points towards keeping, removing or enhancing aspects of a product for future versions.

As an end note to the theoretical framework, it is appropriate to refer to the structure of the remainder of the research. Coherent answering of the two sub-research questions calls for taking a slightly unconventional approach here: data, methodology and results will each appear twice, in two separate but consequential umbrella sections. The first focuses on H1, i.e. establishing the existence of the incongruence, while the second focuses on H2.1 and H2.2, i.e. exploring the incongruence. As such, the feasibility of undertaking the second part can only be demonstrated by completing the first part. In other words, the human judgement approach taken to establish the incongruence, justifies the relevance of employing machine learning in order to explore said incongruence between video game product descriptions and customer reviews. In order to avoid confusion and present a coherent line of reasoning these two parts exist in different sections, each separated into their respective data, methodology and results parts.

### 3. ESTABLISHING THE INCONGRUENCE

This section tackles H1 and the left hand side part of Figure 2.3. Thus, it is concerned with looking at product descriptions and their respective reviews in order to establish a incongruence between these two entities. An incongruence, i.e. differences in content, would point towards a need for reassessing the way descriptions are written to appeal to customers as well as possibly new input for future product development. Importantly, this section makes use of human judges to evaluate the content of descriptions and reviews. It is referred to as the human judgement approach. First, the data collection is described, then the methodology is introduced and lastly, the results are presented.

## 1. DATA

To establish the extent and characteristics of the incongruence between description and review content, many products are required. The products used for this process must be in the same product category for reasonable conclusions to be drawn. A data set which features this category information was published by Julian McAuley (McAuley et al., 2015). It includes customer reviews and category information for these products. Additionally, the respective product descriptions are scraped from Amazon.

The human judgement approach employs a small review-to-description ratio. Therefore, a large corpus of reviews is used (McAuley, 2015). This corpus was mined from Amazon over a 18-year period, between 1996 and 2014. The entire corpus includes over 140 million unique reviews. From this plethora of information, a product group and from within that, a category are selected. Then, the descriptions corresponding to products fitting these criteria are scraped from Amazon. These product descriptions form the basis for human judgement approach part one. It consists of eliciting attributes and consequences common among the product descriptions and tagging them accordingly. It is carried out by a single human judge (the researcher). Subsequently, the reviews corresponding to these descriptions are assessed with the goal of finding: do they include the same attributes and consequences as their respective description? Essentially, this part matches found attributes and consequences to reviews and descriptions and then compares whether the pattern of feature presence in product descriptions is comparable with that in customer reviews. This includes a survey on Amazon Web Services’ (AWS) Mechanical Turk (MTurk), referred to as Amazon MTurk from hereon after. Considering the aim and scope of this section, the review-to-product ratio is small. This ensures that as many products as possible can be assessed within the product category, while limiting the workload for the human judges. In the following subsections, selection of product group and product category, data collection, sampling and the finding of product attributes and consequences are described.

### 1.1 Selection of Product Group and Category

As stated by the researchers who crawled Amazon to gather the data, there are 82.83 million unique reviews after rigorously removing duplicates. There are 24 product groups ranging from books to garden furniture. This research focuses on the product group of video games. Besides the fact that it includes a large amount of products, video games is chosen because it combines two interesting features. Firstly, video games are technology-related, which implies that there might be mentions of factual product specifics such as speed of game, loading time etc. Secondly, video games are often centered around unique game play scenarios, thus adding a more whimsical aspect to descriptions and reviews. This combination of technical, rational and on the other hand narrative, emotional aspects under the umbrella of video games renders this product group attractive for this research.

Within video games category on Amazon it is necessary to select a homogeneous and ubiqui-

tous product category, to ensure comparability of the products. There are 219 of such categories within video games. Assessing them shows, that not all are purely games, but many are related to accessories and additional gadgets. Comparing the attributes and consequences of a gaming console accessory cable to those of a children’s video game would be unreasonable. Intuitively, this is because the former has first and foremost technical characteristics (i.e. it needs to perform an action) and the latter needs to fulfill a child’s desire to experience joy.

In order to arrive at a choice that takes into account the needed ubiquity of the product category, the number of occurrences of all categories is compared. Among the top product categories within video games, gaming consoles form a considerable portion. Yet, even among consoles there are relevant differences affecting the types of games produced for them. This is owed to for example different screen sizes of the medium the game is displayed on or the use of hand-held controllers versus the use of a mouse or the case where control buttons are integrated into the same device as the screen. Fortunately, the versions of *Xbox* and *PlayStation* are widely represented in the data set. Moreover, they are comparable in terms of gameplay and widely accepted to be direct competitors, as exemplified by a number of articles comparing their performance. In fact, they are both played with one controller per player, held in both hands (unlike Wii) and the games are usually displayed on a large TV screen rather than a computer desktop (which is the case for PC games) or a handheld device (Nintendo DS). The combination of these three factors leads to relatively comparable gaming experiences for generations of *Xbox* and *PlayStation*, since for example Wii games are often sport related as their controller can for example be used as a tennis racket.

Thus, the focus is on video games in combination with the different generations of the aforementioned consoles, namely: *PlayStation*, *PlayStation 2*, *PlayStation 3*, *Xbox* and *Xbox 360*. For these products, the corresponding descriptions are scraped from Amazon using the product code (referred to as asin code) supplied as part of the data set.

Sometimes, video games labeled as belonging to the product category of a console are also labeled as belonging to a non-game category. This can be seen looking at category belongings as well as product descriptions. Consequently, products from these categories are excluded. These are *Accessories*, *Hardware*, *Subscription Cards* and *Interactive Gaming Figures*. The final console product frequencies can be observed in Table 1. In total, there are 922 products. For each of these products, there are 5 reviews. There are 5 reviews per product for two reasons. Firstly, this is the highest number of reviews common across all products. Secondly, this ensures a realistic workload for the human judges.

## 1.2 Sampling Procedure

To reduce the data set at hand to a size that is realistic for human judges to assess within reasonable time and monetary expenses, a sample of the relevant products is taken. Since gaming consoles are

**Table 1:** *Counts of video game product descriptions for the PlayStation and Xbox consoles.*

Console Type	Counts	Total per Console Type
PlayStation	90	549
PlayStation 2	207	
PlayStation 3	252	
Xbox	91	373
Xbox 360	282	

a uniting factor in this research, sampling is done with regard to the whole set of products, rather than the separate console subsets. Thus, a random sample without replacement of 200 products is taken (with each 5 reviews, thus 1000 corresponding reviews). This results in a similar allocation among the consoles. This sample of 200 is used throughout this entire section because the goal is to compare the content of product descriptions and their respective reviews. Such a comparison is only sensible when executed on pairs of reviews and descriptions.

### 1.3 Eliciting Product Attributes and Product Consequences from Descriptions

The following describes how the means-end theory is implemented to inspect product descriptions. Gutman’s approach to establishing attributes and consequences connected to a product involves interviewing respondents using the laddering approach. In this research, these are not elicited using interviews, but from the 200 product descriptions.

The researcher elicits attributes and consequences from these descriptions herself. It entails reading all 200 product descriptions and naming the attributes and consequences mentioned in the descriptions. Reading the first descriptions, numerous new attributes and consequences are found. Logically, while reading the later descriptions most attributes and consequences present in video game descriptions have present in an earlier review. So, relative to each description, fewer and fewer new attributes and consequences were elicited the more descriptions are read. Similarly, certain attributes and consequences are grouped together during the reading process. Thus, slightly differently-phrased attributes or consequences that have intuitively similar meanings, such as *visuals* and *graphics* are combined into one, e.g. *graphics*. Another example for this is *fun*, *exhilarating* and *exciting*, which are all grouped into *exciting*. Lastly, the video game descriptions often circle around the narrative content of the games, which especially include mentions of the *storyline* and *characters*. These attributes are marked as present even if the description does not explicitly mention the words "character" or "storyline" and was rather describing the story or personalities present in it. This is because it can be expected from human judges to make the connection between flowing text and its topic (e.g. *storyline*). Similarly, the attribute *team manager* is applied to mostly sports-related games where part of the task is to lead a team to success. Here, also the consequence *prove skills* is then present because in order to make a team strong, the player would have to prove their skills.

One could argue that this is a generalization of discrete attributes and consequences. However, it is done in full awareness. An example on alcoholic drinks of a means-end laddering chain provided in Reynold and Gutman’s 1988 article in the Journal of Advertising Research, illustrates this. The authors state their reasoning to summarize a list of negative consequences of alcohol consumption as *avoids the negatives of alcohol* elicited during their interview process as being that the lack of generalization would lead to low frequency of repetition in the data and thus low frequency of matches.

Simultaneously, a matrix relating the overall found attributes and consequences is gradually built, with attributes denoting columns and consequences denoting rows. The complete set of attributes and consequence and values is seen in Table 2. The crosses represent related attributes and consequences for easier understanding.

Below is an example of a product description, with the asin code B0050SWTS8. In the following, attributes are highlighted in pink and consequences highlighted in green. The name of the attribute or consequence in the text is mentioned in square brackets and italics following the relevant part of the text.

Dead Space 3 brings Isaac Clarke and merciless soldier John Carver [*character*] on a journey across space to discover the source of the Necromorph outbreak. Crash-landed on the frozen planet of Tau Volantis, Isaac must comb the harsh environment for raw materials and scavenged parts. He will then put his engineering skills to the ultimate test to create and customize [*customization*] weapons and survival tools. The ice planet holds the key to ending the Necromorph plague forever, but first Isaac must overcome [*prove skills*] avalanches, treacherous ice-climbs, and the violent wilderness. Facing deadlier evolved enemies and the brutal elements, Isaac can choose to team up [*interactive*], not only for his own survival, but for that of mankind [*save the world & responsibility*]. Play together with a friend or go it alone [*single and multiplayer*] as Isaac Clarke using the seamless new drop in, drop out co-op functionality. Each mode [*modes of play*] offers unique story [*storyline*] elements and gameplay.

Then, all descriptions are read again, and they are tagged according to whether an attribute and a consequence are present in it. The output of the description tagging process is as a binary matrix with 200 rows representing product descriptions and 18 columns representing attributes and consequences.

**Table 2:** *Attributes, consequences and respective values present in product descriptions of video games, elicited using human judgment.*

Attributes	Consequences										Values
	Exciting	Interactive	Realistic	Control over game	Save world	Prove skills	Identification	Responsibility	Creativity	Nostalgia	
Modes of play		x									
Single and multiplayer		x				x					
Storyline	x		x		x			x			
Sound	x		x				x			x	
Character	x						x	x			
Graphics			x				x			x	
Team manager				x		x		x			
Customization		x		x			x		x		
Fun and Enjoyment in Life	Being Well Respected	Being Well Respected	Sense of Belonging	Sense of Accomplishment	Being Well Respected	Sense of Accomplishment	Self-Fulfillment	Being Well Respected	Self-Fulfillment	Sense of Belonging	

## 1.4 Matching Product Attributes and Product Consequences with Reviews

The goal of this subsection is to tag all reviews according to whether the attributes and consequences found previously are present in them or not. As there are 200 tagged descriptions, and there are 5 reviews per description, 1000 reviews are used in this process. These reviews are presented to workers on Amazon Mturk. The details of the process are described in the following. MTurk is a platform ran by Amazon where it is possible to register either as a worker to complete tasks or as a requester to publish tasks. A task is referred to as a Human Intelligence Task or HIT. In this research, the task for the workers is to, for each review, classify which of the product attributes and consequences elicited using the means-end chain on video game descriptions are present.

The incentive for workers to join MTurk’s workforce is the salary earned per completed HIT. For the requester, the benefit of MTurk lies within the access to a large and active workforce, enabling a swift gathering of data. As in any employer/ employee relationship, there are factors that must be taken into account before setting a salary and designing the HIT. The salary per completed HIT should be in fair relation to the time spent on completing the HIT. Otherwise, workers might either hand in lower quality work due to frustration. On the other hand, the larger the salary per completed HIT, the larger the costs to the researcher. Thus, there is a implicit trade off between quality of work and price to pay.

Appendix A shows the HIT as presented to workers on Amazon MTurk. In light of the relatively complex instructions featured in the HIT for this research, each HIT contains all 5 reviews per product. Differently from a lot of HITs that feature easy one-sentence instructions, this HIT requests more attention to the longer instructions. The hope is that economies of scale lead to a lower total time needed for all workers to complete all tasks, and thus a lower total cost to the researcher. Furthermore, a learning curve is assumed. Finally, the reward per completed HIT was set to be equal to 0.40\$. As an extra step to ensure a certain quality of work regardless of the payment/time ratio, MTurk enables requesters to specify what kind of workers should be allowed to work on the HIT, by offering qualifications. The qualifications in the case of this research’s HIT are as follows:

- Location *is* UNITED STATES (US)
- Number of HITs Approved had to be *greater than* 1000
- HIT Approval Rate (%) for all Requesters’ HITs had to be *greater than* 95

The following illustrates a case of analyzing a full set of five reviews as expected to be fulfilled by workers on MTurk. The product used for this is B0050SWTS8, as in the explanation of the laddering approach using its description in subsection 1.3 (Eliciting Product Attributes and Product Consequences from Descriptions). The goal is to show mismatches and present attributes and consequences. Per review, workers select which of the 18 product attributes and consequences are

present. It must be noted that the review texts are not pre-processed for grammatical mistakes or misspellings, in order to present the workers on MTurk with as realistic a situation as possible. Attributes are highlighted in pink and consequences highlighted in green. The output of the HIT and thus this subsection is a binary matrix of 1000 rows and 18 columns. The index consist of asin codes, each repeated 5 times. The columns are attributes and consequences.

#### Review Text 1

I like this one as much as the other ones, the gore **gore** [*mature*] is great. I would recomend this game to all

#### Review Text 2

While this game isn't perfect it is WHY we need **single player** [*single and multiplayer*] campaigns, **Solving puzzles** [*prove skills*], cracking codes, reassembling aliens ! None of this stuff is ever done on **multi-player** [*single and multiplayer*]. This game is a good template for FPS single player missions to come.

#### Review Text 3

I love this game, and nearly all of it is fine, but for me, the door locks that **require unlocking** [*prove skills*] before you can go into the next room have become to difficult for me to master, and 60% into the game I am stuck since I am way to lousy of a player to proceed past the locks. other than that, great!!

#### Review Text 4

Dead Space three was a lot of **fun** [*exciting*] to play. I played the co-op with my best friend and had a good time with it. This game is quite a bit different from the first two Dead Space games. It's not quite as eerie or scary as the previous installments. I realize that this comes largely from playing **side by side with a friend** [*single and multiplayer*] in co-op **mode** [*modes of play*]. **Single player** [*single and multiplayer*] still offers a few good jumps here and there. As far as the **story** [*storyline*] goes...it is pretty convoluted and honestly I don't even remember much about it (it's not all that memorable). Still though, it is fun customizing your gear and running around **gunning down necromorphs** [*mature*]. The guns you can make are really cool, except it seemed like (in my opinion) there were only a few good gun combinations. I do like how they added the scavenge and **customization** [*customization*] options though. Anyway, I think overall it was a really good game but I don't think it quite deserves 5 stars.



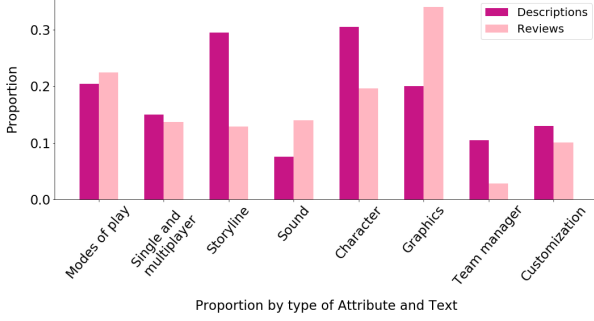
4 stars because I with is wasn't an obvious full port and was not as scary [*exciting*] to play on a big screen TV with a nice surround system [*sound*] as the last one

## 1.5 Data Description

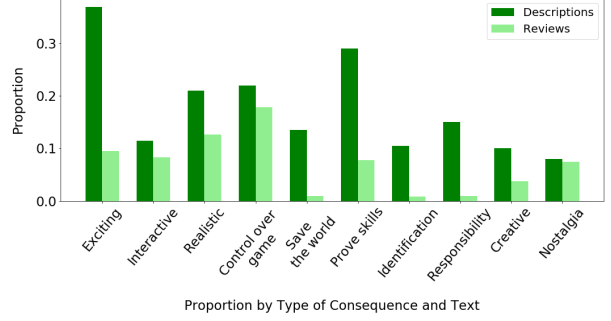
This part describes the data collected in the previous two subsections: 1.3 and 1.4. The former presents observations which are descriptions tagged with respect to which attributes and consequences are deemed present in them. The latter presents observations which are reviews tagged with respect to which attributes and consequences are present in them. Firstly, this research looks at the prevalence of the elicited attributes and consequences. During data collection presence or absence of an attribute or consequences was indicated by a 1 or 0, respectively. Therefore, aggregates show the prevalence of attributes or consequences in the corpora. Before proceeding, all attributes and consequences found are evaluated with respect to their prevalence overall. This is done on the description findings, because they feed into the review data. The consequence *mature* is the only parameter recorded as present in less than 5% of the total number of descriptions. Therefore, it is removed for further analysis. Also, it can be assumed that if a game has mature content, that would be signified in form of age restrictions on the game packaging. Additionally, labeling a game as mature in descriptions could deter e.g. parents from buying a game for their children.

As mentioned before, 200 video game descriptions are sampled. For each of these, 5 reviews are used to complete the HIT on MTurk, resulting in a total of 1000 reviews. Each description corresponds to five reviews. Thus, there are five description - review links per product. To compare the relative prevalence of elicited attributes and consequences among the two groups of texts examined, the proportion of texts classified as mentioning these is calculated. The column sums of the output matrix of assigning attributes and consequences to product descriptions. It depicts which descriptions mentioned which attributes and consequences, is divided by 200. The column sums of the output matrix of which reviews mentioned which attributes and consequences, is divided by 1000. These denominators represent the maximum sum per column, thus these calculations result in the proportion of texts mentioning the terms in question.

Figure 3.1 and Figure 3.2 depict the above mentioned proportions. As seen in Figure 3.1 for product attributes, the most notable discrepancies between proportionate presences in descriptions and reviews is seen for *storyline*, *sound*, *character*, *graphics* and *team manager*. *Storyline*, *character* and *team manager* are relatively more present among descriptions. Both *storyline* and *team manager* are attributes that are more than twice as prevalent among descriptions than among reviews. On the other hand, *sound* and *graphics* are relatively more prevalent in reviews than in description texts. *Sound* is almost twice as prevalent in review texts than in descriptions. *Graphics* is roughly one third more prevalent in review texts. Among all attributes examined, the ones



**Figure 3.1:** Bar chart showing the prevalence of elicited attributes among the sampled descriptions and respective reviews



**Figure 3.2:** Bar chart showing the prevalence of elicited consequences among the sampled descriptions and respective reviews

which are proportionally most similarly represented in descriptions and reviews are: *modes of play*, *single and multiplayer* and *customization*.

Regarding product consequences, the main observation is that all consequences are perceived to be relatively more prevalent in descriptions than reviews. The lowest discrepancies between proportions are observed for the following consequences: *interactive*, *realistic*, *control over game* and *nostalgia*. Conversely, especially *save the world*, *identification* and *responsibility* are almost nonexistent in the perceptions of review content. Moreover, *prove skills* and *exciting*, the two proportionally most prevalent consequences in description texts, are perceived to be less than a third as prevalent among review texts.

## 2. METHODOLOGY - MULTIPLE CORRESPONDENCE ANALYSIS

When aiming to uncover latent pattern in big datasets with a high number variables, dimension reduction techniques are a useful measure to identify the most relevant dimensions. A solution to this is Multiple Correspondence Analysis (MCA). It is similar to Principal Component Analysis (PCA), except for the type of variables they are applied on. Whereas PCA uses continuous variables, MCA is appropriate for the analysis of categorical variables. The goal of MCA is to reveal meaningful relationships between categorical variables (Hoffman & De Leeuw, 1992). As a result of employing MCA the most informative dimensions can be plotted on a lower Euclidean space to visualise relationships between variables. The most informative dimensions are those that explain the most variance. Such visualisation of results is one way to arrive at meaningful interpretations from MCA.

MCA is also known as homogeneity analysis since it attempts to group similar observations. Similarity is established based on the centroid principle which assesses the closeness of e.g. variable category points. Using this principle, the variable category points are plotted in the center of observations that belong to that category level. As such, it is implied that MCA plots observations that belong to the same category level relatively close to each other on the Euclidean plane, while

observations that do not occur in the same category are plotted relatively far from one another. Since observations that are part of the same categories will be relatively close, it follows that categories that share similar observations are close.

FOR MCA, assume a data set of  $n$  observations for  $m$  categorical variables.  $p$  being the number of dimensions, the data set is displayed as an  $n \times p$  matrix  $\mathbf{X}$  of observation scores. Let

- $K_j$  be the categories of variable  $j$ ,
- $\mathbf{G}_j$  be an  $n \times K_j$  indicator matrix and
- $\mathbf{Y}_j$  be the  $K_j \times p$  matrix of category coordinates for variable  $j$ .

In  $\mathbf{G}_j$ , every row sums to 1, because each row represents for the respective observation which category of variable  $j$  it belongs to.  $\mathbf{G}_j \mathbf{Y}_j$  is an  $n \times p$  matrix. In row  $i$ , for category  $k$ , matrix  $\mathbf{G}_j \mathbf{Y}_j$  contains the category coordinates if that  $i$ th observation occurs in  $k$ .

Categories are fit to the dimensions using alternating least squares; the aim is to reduce the distance between category and dimension to a minimum. How well a variable is explained by and this discriminates along a dimension is examined by looking at so-called discrimination measures. For variable  $j$ , the discrimination measure is a squared correlation between  $0 \leq d_{js} \leq 1$ . If a variable has a higher discrimination measure, it implies a better fit of that variable on the dimension in question. The arithmetic mean of all discrimination measures within a dimension is that dimension's eigenvalue. That eigenvalue indicates the variance of the data explained in the respective dimension. The higher the eigenvalue of a dimension, the more meaningful and interpretable the respective dimension. For MCA, a commonly used correction for eigenvalues is the Greenacre correction (Abdi & Valentin, 2007). It takes into account that for MCA, each category level is represented by two binary columns, representing category belonging or absence for each observation. That is, the number of columns is twice that of the categorical variables. The row sums, are always equal to the number of categories.

When determining the number of dimensions that qualify as suitable to be included in analysis, a screeplot can be used. It represents the magnitude of the eigenvalue (variance explained) per dimension. These points on the graph can be connected with a line for better interpretability, since the usual cutoff point is made at the elbow of the graph. That is the first dimension after a sharp decrease in eigenvalue, is the first not to be looked at. It looks like a flattening of the connections between points. This is because this levelling off of the eigenvalues represents a stagnation in added explanatory power of the dimensions with said low eigenvalues.

When gathering data, the collected data is not the entire population, but is merely an observed part thereof, a sample. The uncertainty that comes with treating a sample as the population, can be taken into consideration using bootstrapping. It falls under resampling techniques. It uses random sampling with replacement; thus, bootstrapped samples are created which are of the same length as

the original data set, but where rows are randomly drawn from the original data set with replacement; individual indices representing the same row from the original data set can appear more than once. Repeating the resampling process numerous times, the resulting bootstrapped samples facilitate the simulation of an expression such as the aforementioned dimension eigenvalue using these. To measure the stability of a data set, the bootstrapped eigenvalues can be used to compute a bootstrapped confidence interval, which is then compared to the population eigenvalue for the respective dimension.

Once the number of dimensions to be visualised has been determined, a joint plot can be used to show two dimensions at a time, each represented by an axis. On the plane, category points are plotted, according to what discrimination measures they have for the respective dimensions. Thus, the rationale behind the above mentioned closeness of category points that include similar observations comes into play again. The farther apart category points belonging to the same variable are spread across the plot, the better that variable discriminates, i.e. the more meaningful its interpretation. The farther away a category point is from the mean, i.e. the origin of the plot, the more interpretable it is, and the better it discriminates. Generally, if two category points are at opposite ends of an axis, it means that observations that are about the one category point are not about the other.

### 3. RESULTS

On the one hand the results concern 200 video game product descriptions (see Subsection 1.3). A sample of the data used can be seen in Appendix C. It is shown, that each row contains the information on one observation, here, one description. Therefore, there are 200 rows. The table indicates the presence or absence of attributes and consequences in the description. The umbrella terms attributes and consequences, represent 18 two-level categorical variables, elicited using the means-end chain. There are 8 attributes and 10 consequences. Each of these has two levels, indicating said presence (1) or absence (0) of the attribute or consequence in question in the observation. The second part of the results concerns 1000 video game customer reviews (see Subsection 1.4). A sample of the data used can be seen in Appendix D. The only difference with respect to the structure of the data compared to the above mentioned data, is, that there are 1000 rows, because here, observations represent reviews.

Multiple Correspondence Analysis is implemented on both data sets separately, to explore and compare the different configurations of the attributes and consequences in the two forms of text: product descriptions and customer reviews. Using Greenacre correction on the description data, the first two dimensions account for a variance explained of 63.69%. On the review data, this amounted to 72.97%. The variance explained corresponds to the eigenvalue of the respective dimension. For both descriptions and reviews, the first two dimensions will be compared. The number of dimensions to be looked at should be those until the flattening out. As seen in Figures 3.3 and 3.4, the elbow of the screeplot is at dimension 3 for descriptions and dimension 2

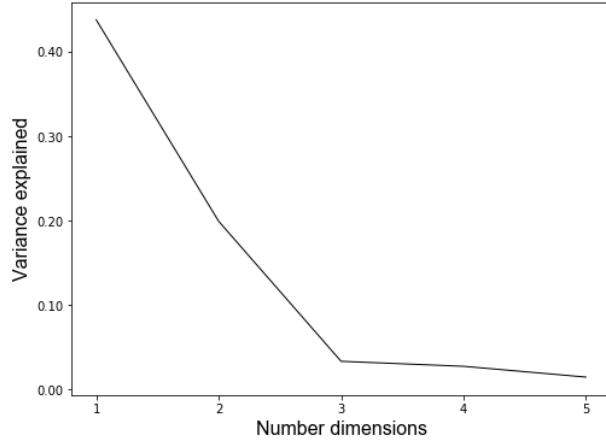
for reviews. In the case of reviews, the elbow is at the second dimension, i.e. that dimension explains much less variance than the first dimension. This would imply usage of only dimension 1 in the plot, however it is easier to visualize the data including the second dimension. Therefore, dimension 2 from the review data will be included in the analysis to create a two-dimensional joint plot.

Each of the categories representing the 2 possible levels of the 18 attributes and consequences is plotted on Figure 3.5 for descriptions and Figure 3.6 for reviews. Presence indicates cases where the attribute or consequence was deemed as being part of the content. Absence indicates cases where the respective attribute or consequence was missing from the text. For the 200 product reviews, it can be seen that the category points are spread out over the plot. However, there is a cluster of category points around the midpoint, consisting mostly of absence-indicating category points. This implies that for most attributes and consequences, the presence category point is farther from the origin, and thus farther from the average. That is, the presence of attributes and consequences can be interpreted as more meaningful than their absence. This applies to those presence-indicating category points that lie farther from the origin than their absence indicating counterparts. The same interpretation can be applied to Figure 3.6, where an even more dense cluster of absence-indicating category points can be observed surrounding the origin. The presence of all attributes and consequences is more meaningful than their absence. As a result of this, it is of interest to investigate the exact spread and location of the presence-indicating category points. This is done separately for descriptions and reviews in the following subsections.

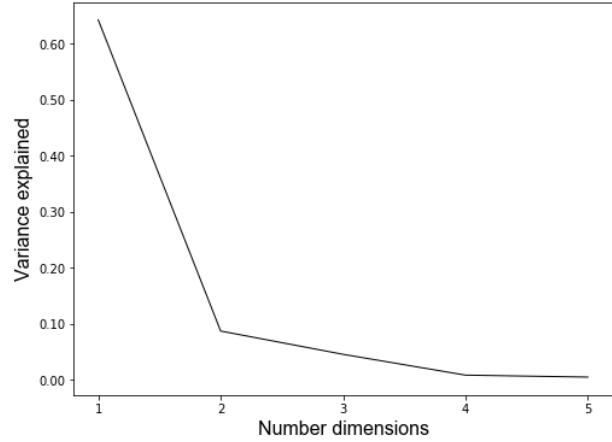
### 3.1 Video Game Product Descriptions

The following describes the results of MCA on product descriptions. The joint plot displays the scores for all attributes' and consequences' categories on the presence level (i.e. the attribute or consequence was present in the description).

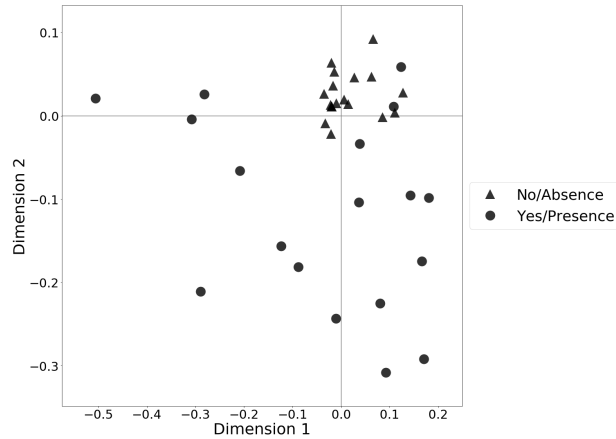
As seen in Table 3, dimension 1 accounted for 43.78% of the variance, using Greenacre correction. Figure 3.7 shows dimension 1 discriminates slightly better on consequences (marked by green) than on attributes (marked by pink). Dimension 2 discriminates better on both attributes and consequences overall, than dimension 1, because as can be seen in especially in quadrants III and IV. In general, there are only two distinguishable clusters of category points that apply to both dimensions. The first one discriminates well on dimension 1 and consists of the attributes *storyline* and *character*. Further category points that score similarly high on dimension 1 are the consequences *save the world* and *responsibility* as well as *identification*. As Table 2 displays, the former two are related to the attribute *storyline*, while the latter is related to the attribute *character*. Another cluster that discriminates well on both dimensions consists of the attribute *sound* and the consequence *realistic*. These are connected in Table 2. As the only category points present in quadrant I, *nostalgia* and *graphics* could be described as a cluster as well, and are displayed as connected in Table 2. Overall, dimension 1 contrasts consequences that are related to the values *being well – respected* and



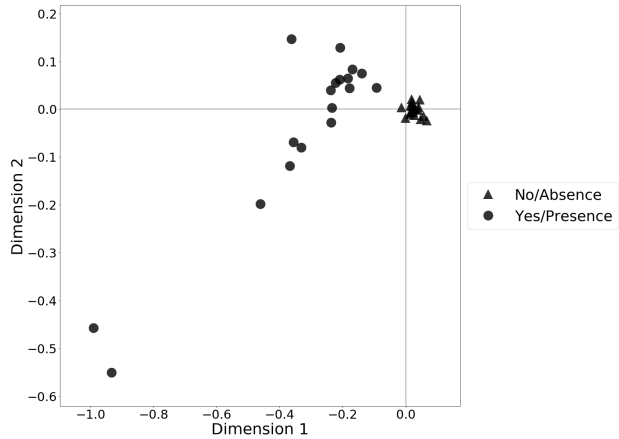
**Figure 3.3:** *Screeplot showing the variance explained for dimensions 1 through 5 for product descriptions*



**Figure 3.4:** *Screeplot showing the variance explained for dimensions 1 through 5 for customer reviews*



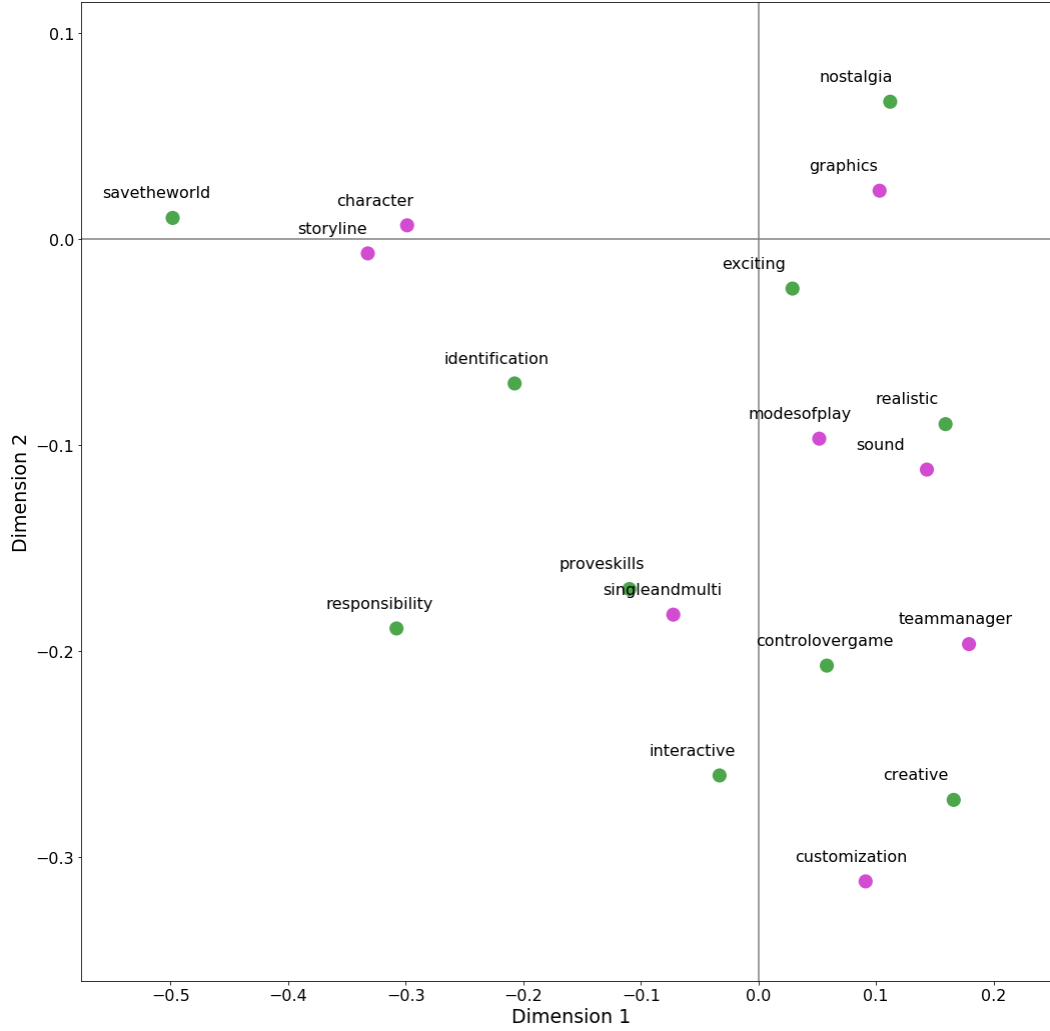
**Figure 3.5:** *Joint plot from MCA on 200 product descriptions showing both category points indicating both the absence as well as presence category points of the respective 18 attributes and consequences*



**Figure 3.6:** *Joint plot from MCA on 1000 customer reviews showing both category points indicating both the absence as well as presence category points of the respective 18 attributes and consequences*

*sense of accomplishment* (*save the world, prove skills, identification and responsibility*) with consequences that are related to the values *sense of belonging* and *self – fulfillment* (*nostalgia, realistic, creativity*). Thus one could say that it discriminates between a sense of doing (fulfillment by actions (*modes of play, storyline*)) and a sense of feeling (fulfillment by sensing: hearing (*sound*), seeing (*graphics*)). The fact the the category points for the attributes *storyline* and *character* on the one side, are on the opposite side of dimension 1 than most of the other attributes (except *single and multi*) indicates that descriptions in which the former is present, are not about the latter.

19.91% of the variance was explained by dimension 2, using Greenacre correction. Dimension 2 appears to discriminate best and similarly strong on the attribute customization as well as the consequences *interactive* and *creative*. These two consequences are connected to the attribute *customization* in Table 2. Dimension 2 also discriminates similarly strong on the attributes *team manager* and *single and multiplayer* as well as the consequences *control over game, responsibility* and *prove skills*. According to Table 2, the former attribute ( *team manager* ) is connected to all three of these consequences, while *single and multiplayer* is merely connected to *prove skills*. Since these category points are closest to the origin, the two variables that the two first dimensions discriminate on least are *graphics* and *exciting*. Looking at only attributes, the Figure 3.7 reveals that *storyline* and *character* are farthest away and thus most different from the sound, *team manager* and customization. *Graphics, modes of play* and *single and multiplayer* are situated slightly closer. Overall dimension 2 seems to contrast functionality-related attributes (*customization, team manager , modes of play , single and multiplayer*) and consequences (*control over game, realistic, creativity, responsibility*) with whimsical, form-related attributes (*graphics, storyline, characters*) and consequences (*nostalgia, save the world, exciting*). In line with Table 2, the former consequences are connected to the values *sense of accomplishment* and *being well respected*. The latter consequences are connected to *sense of belonging* and *self – fulfillment*. One could say that, here, dimension 2 discriminates between function and form. Similarly, the category points for *customization, creative* and *interactive* lie on one end of dimension 2 axis, while *nostalgia* and *graphics* are on the other end. This implies that descriptions that are about the former, do not include mentions of the latter.



**Figure 3.7:** Joint plot from MCA on the 200 product descriptions displaying only the category points indicating presence of the 18 attributes (pink) and consequences (green)

### 3.2 Video Game Customer Reviews

After removing the negatively scoring category points for review data the first two dimensions reveal a plot as shown in Figure 3.8. Immediately, one notices the difference to Figure 3.7. There, category points are spread relatively evenly apart, while in Figure 3.8 the points are clearly clustered. In Figure 3.8, dimension 1 discriminates better among both attributes (marked by pink) and consequences (marked by green) than dimension 2. In both these dimensions overall, consequences are better discriminated against than attributes. The only two category point that clearly form a cluster far from the origin in both dimensions are the consequences *identification* and *responsibility*. They are related with one another according to Table 2 via the attribute *character*. Here, however, *character* lies far closer to the origin, suggesting less discriminatory relevance.

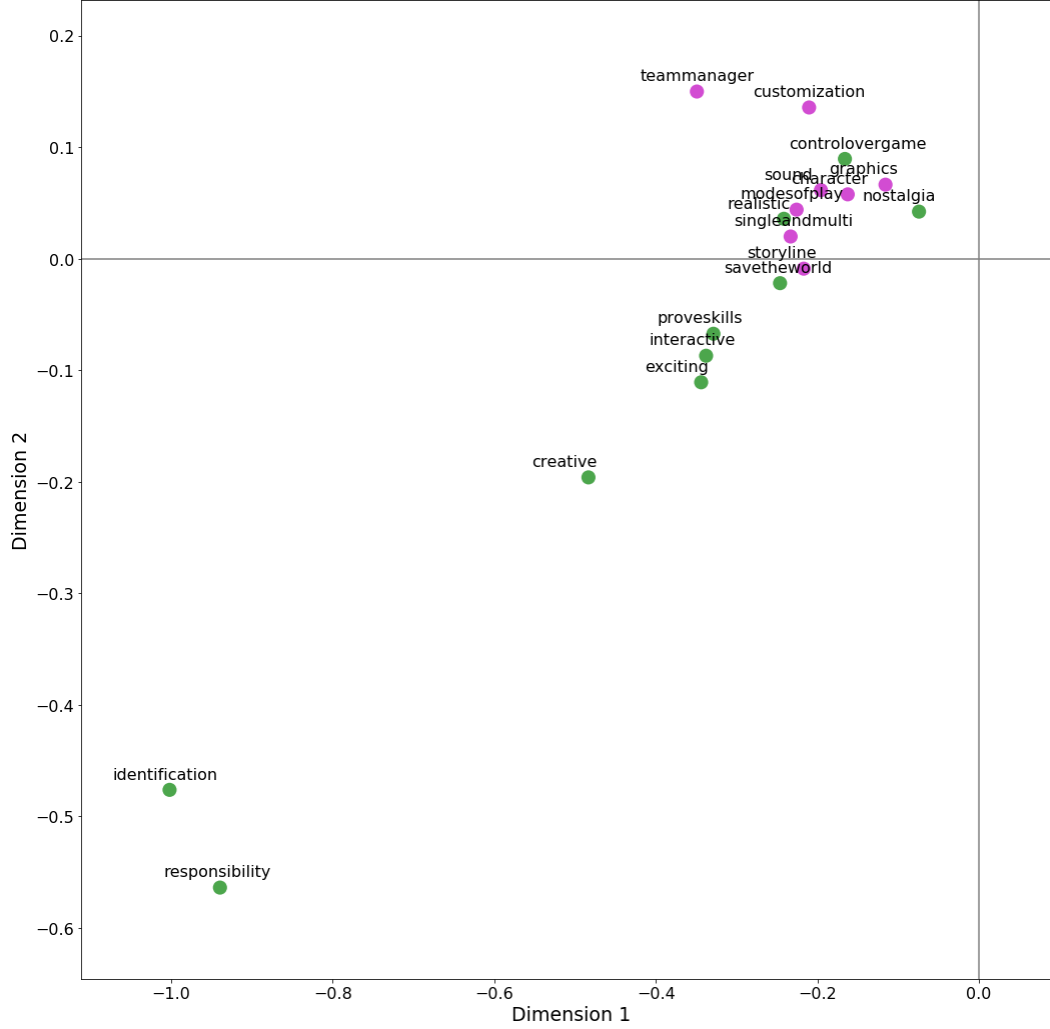
As seen on the right hand side of Table 3, 64.29% of the variance was explained by dimension 1,



using Greenacre correction. Dimension 2 accounted for 8.68% of the variance, using Greenacre correction; considerably lower than the variance accounted for by dimension 1. Closer to the origin from the perspective of dimension 2, but distinctly removed according to dimension 1 is a cluster consisting of the consequences *prove skills*, *interactive* and *exciting*. The attribute *team manager* discriminates similarly well on dimension 1. However, the attribute *team manager* leads to the consequence *prove skills* in Table 2. On both dimensions, the attribute *character* and the consequence *control over game* score equally, however, they are not connected via Table 2. *Creative* discriminates well on both dimensions, but is not close to its respective attribute *customization*. Close to the origin with respect to both dimensions, lie most attributes. *Storyline* and the consequence *save the world*, also connected on Table 2, score similarly on both dimensions but low on dimension 2.

On dimension 1 especially, the consequence *realistic* is closely situated to two attributes connected to it: *storyline* and *sound*. The attributes *customization* and the consequence *control over game* are discriminate similarly on dimension 1 and are connected on Table 2. On both dimensions, *graphics* and *exciting* have the least meaningful discriminatory scores. However, on dimension 1 *graphics* is the closest attribute to *nostalgia*, corresponding to Table 2. Looking at the overall spread of category points across dimension one, it is found that *identification* and *responsibility* seem to be in customer reviews that are not about *nostalgia* and *graphics*. However, it has to be said that the latter two are rather close to the origin, indicating that they do not express a lot about reviews. Among attributes, it appears that reviews that are about *team manager* are not about *graphics* and *storyline*. As said before, category points far away from each other can be interpreted to appear in different kinds of texts. However, it must be kept in mind that the variance explained by certain dimensions is lower, as with the case of customer reviews, where dimension 2 displays a much lower population eigenvalue than dimension 1.

Overall dimension 1 seems to contrast possibly modern, action - related attributes (*team manager*, *modes of play*, *single and multiplayer*) and consequences (*creative*, *interactive*, *identification*, *prove skills*) on the one hand, with more traditional and playful attributes (*graphics*, *sound*) and consequences (*nostalgia*) on the other. The dimension appears to be less content-related, as *storyline*, *character* and *realistic* appear closest to the average discrimination scores of this dimension. The value most clearly connected to the action-related side of dimension 1 is *self – fulfillment*. The other, more playful side of this dimension corresponds to the value *sense of belonging*. Dimension 2 on the other hand, seemed to contrast a reviews with a sense of gamer singularity (ego-focused) versus those with a view on *interaction* with others (togetherness). For the former - "ego focused" - this is exemplified by the value belonging to the strongly discriminated consequences (*identification*, *creativity*): *self – fulfillment*. For the latter - "togetherness" - the consequences (*realistic*, *nostalgia*, *control over game*, *prove skills*, *interaction* and *save the world*) were related to the following values: *sense of belonging*, *sense of accomplishment* and *being well – respected*.



**Figure 3.8:** Joint plot from MCA on the 1000 customer reviews displaying only the category points indicating presence of the 18 attributes (pink) and consequences (green)

### 3.3 Testing the Stability of the Data by Observing Bootstrapped Dimension Eigenvalues

In order to test the stability of the description and the review data set, the bootstrapping technique is used. This resampling technique can be used to estimate the validity of results. It can estimate measures of accuracy of bootstrapped samples, for example using confidence intervals and then comparing that bootstrapped confidence interval to the observed population parameter. In this research, bootstrapping is done as described in the following. The product description data set (see Appendix C) will be used to create 1000 bootstrapped samples, and then computing the bootstrapped 99% confidence interval for eigenvalues per dimension. Random resampling with replacement is applied to the observed description data set 1000 times resulting in 1000 bootstrapped samples of the same number of observations as the original data set, here 200 rows of observations. For each of these resamples, estimate sample statistics are computed here, the eigenvalue per dimen-

**Table 3:** *Population eigenvalues versus bootstrap confidence intervals.*

	Product Descriptions		Customer Reviews	
	Population	99% CI	Population	99% CI
Dimension 1	0.4378	[0.3142, 0.3222]	0.6429	[0.6808, 0.6868]
Dimension 2	0.1991	[0.1736, 0.1787]	0.0868	[0.0841, 0.0872]
Dimension 3	0.0331	[0.1117, 0.1143]	0.0452	[0.0355, 0.0371]
Dimension 4	0.0272	[0.0625, 0.0649]	0.0080	[0.0128, 0.0134]
Dimension 5	0.0144	[0.0156, 0.0164]	0.0045	[0.0034, 0.0036]

sion. It indicates the variance accounted for by a dimension. Using these 1000 dimension-specific eigenvalues, a confidence interval for this dimension's eigenvalue distribution is computed. The observed population eigenvalue is then compared to this bootstrapped confidence interval. This is used to test the stability of the results. If the observed value lies outside of the confidence interval, it cannot be assumed with 99% certainty that the population eigenvalues lie within the bootstrapped eigenvalue confidence interval. This indicates that the data set is not stable, negatively impacting how representative and replicable the original data set is.

For both the product description data set (see Appendix C) and the customer review data set (see Appendix D), bootstrap resampling is repeated 1000 times each. For both, the sample statistic chosen to compute the 99% confidence interval are the dimension eigenvalues. The bootstrapped eigenvalues are computed for the entirety of the respective bootstrapped samples for the first five dimensions for each data set. Using these bootstrapped eigenvalues, 99% confidence intervals are constructed, one per dimension. The results of this are reported in Table 3. For product descriptions, Table 3 shows that the population eigenvalues for none of the first five dimensions lie within the bootstrap 99% confidence intervals. Therefore, for the descriptions data set used in the analysis of this section, it cannot be assumed that 99% of the intervals from the bootstrap samples include the population eigenvalue. For customer reviews, Table 3 shows that only the eigenvalue for dimension 2 is within the 99% confidence interval for that dimension. Thus, it has to be concluded that neither of the data sets in questions can be considered stable. Rather, it must be said that the samples are likely both unrepresentative and display low replicability. This impacts the reliability of the results. Because the analysis in this part of this research is explorative and it consisted to a large extent of attempting to formulate an alternative to conducting interviews, finding the data is not stable is not detrimental. It is known that several limitations during the collection of data were present: limit of financial resources, time and collection of data on Amazon MTurk.

### 3.4 So what? Implications from the Existence of the Incongruence

When comparing the joint plots displaying the first two dimensions of MCA for descriptions and reviews, the structure and patterns observed are undoubtedly different. Figure 3.7 displays evenly

spread out category points with two small exceptions. In contrast, Figure 3.8 exhibits highly clustered category points, with few dispersed outliers. It follows that the interpretations of the dimensions of the two figures vary between descriptions and reviews. In the case of product descriptions, the first dimension discriminates between description texts that contain action-related variables and those that are concerned with visual and auditory aspects of video games. Dimension 2 in video game descriptions discriminates between functional and formal features. With respect to video game reviews, the first dimension contrasts those reviews focusing on contemporary upgrades and modern functionalities with those which contain information on traditional and aesthetic aspects. The second MCA dimension on review texts diversifies between reviews focused on address the views of a gamer as an individual and reviews that contain information on togetherness of the player with other gamers. As a result,

H1: The product attributes and product consequences observed in video games display a different pattern in product descriptions than in customer reviews.

cannot be rejected. The presence, structure and prevalence among attributes and consequences differs observably between the two data sets. Thus, it is of value to conduct a study on the extent of discrepancies between review texts and respective product descriptions. The managerial relevance of this exists in regard to the content of descriptions with their goals of informing and confirming a customer's thought process in relation to purchases. It is thus of interest to inspect what the possible reasons for the discrepancy shown by MCA between category prevalence and patterns in descriptions and reviews are. The next section of this research will focus on exploring the incongruence that was established among descriptions and reviews for the video game category. The following section is thus dedicated to a single video game with a large set of reviews. This ratio is also more representative of a real business situation as managers will most likely focus on few products and require video game-specific product information on an individual rather than on an aggregate level as discussed in this results section. Nonetheless, the now ending section laid the groundwork necessary to justify a deeper exploration of the incongruence.

## 4. EXPLORING THE INCONGRUENCE

This section tackles answering hypotheses 2.1 and 2.2 and the right hand side part of Figure 2.3. Thus, it is concerned with looking at customer reviews of a single product. The aim is to identify what the established incongruence in section 3 looks like on the product level and to formulate possible improvements to a description based on a single case. This section does not make use of human judges to inspect the reviews, but rather natural language processing, therefore, it is referred to as the machine learning approach. First, the data collection is described, followed by the methodology and the presentation of the results.

## 1. DATA

The machine learning approach focuses on eliciting attributes from reviews, in order to facilitate improvement of the respective product descriptions, as seen in the right hand side part of Figure 2.3. For this, a large review-to-description ratio is needed. The data set used in the section 3 of this research is not appropriate, since it features small review-to-description ratios. Instead, a data set provided by Amazon Web Services is used ("Amazon Reviews pd", n.d.). It features a list of links to review files on a range of product groups on Amazon and has over 1.7 million entries of customer reviews for the video game product group. The data set does not have product descriptions. Moreover, there is no information on product categories, thus this data set would not have been suitable for the human judgement approach. Within this data set, again, only actual video games for the consoles established in 1.1 are eligible for further analysis in this research.

### 1.1 Selection of Product to Examine

The product for this part of the analysis is chosen based on several conditions. Firstly, there has to be a sufficient amount of reviews available. Secondly, the product has to be a video game, and not an accessory or similar (see section 1.1). Thirdly, the proportions of positive and negatively-rating reviews have to be balanced. These ratings are assigned to the game by the author of the respective review and appears together with it. Reviews that gave 4 or 5 stars are considered to be positive and reviews that gave 1 to 3 stars are considered to be negative, so there is a division according to these two parts. A balanced set of reviews implies a game that was reviewed with mixed ratings, i.e. has similar proportions of positively versus negatively-rating reviews. The reason for the need for balanced data is as follows. Intuitively, it can be suspected that positively-rating reviews mention different aspects of the game than negatively-rating reviews, as this would imply that there are generally favored features of the game and generally disliked features. Using these conditions, the following video game is selected: *Call of Duty: Ghosts* is selected with 599 positive and 374 negatively-rating reviews. *Call of Duty* is a ego-shooter game where the player has to kill opponents. It can be played either online against other players or in single player mode, where a storyline is followed, called campaign. There are yearly releases of new games, with technical improvements and different storylines. The development of the game is alternatingly done by two production firms: Activision and Treyarch. The version of the franchise central to this section, *Call of Duty: Ghosts*, was developed by Activision. In Figure 4.1 based on the aggregate of reviews having received one of the five possible star ratings, the proportion of reviews with one such rating out of all video game reviews is displayed. For ratings of 1 to 3, *Call of Duty: Ghosts* got relatively more than video game reviews at large. For ratings 4 and 5, it has relatively less. This illustrates that *Call of Duty: Ghosts* overall was rated more critically than products in the video game category at large on Amazon.

### 1.2 Pre-Processing

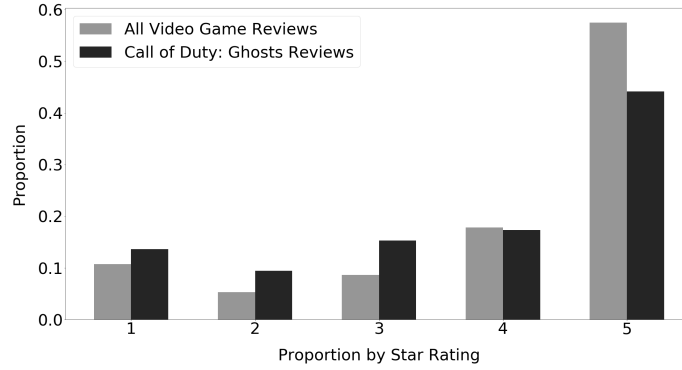
Textual data has to be pre-processed. In this research this consists of the steps described in the following. Firstly, characters and quotes are removed. Then, all sentences are transformed into

words. Consequently, bigrams and trigrams are filtered out. The thresholds for these are set to 50 occurrences. The threshold ensures that only ngrams that occur often are taken into account. Then all words are lemmatized, i.e. reduced to their inflectional form, so that e.g. differences in tense can be ignored. After lemmatization, stop words are removed. A list of words was added to the general list of stop words because these words occurred often in the texts but did not have any expressive value: 'video', 'game', 'great', 'good', 'cod', 'call', 'duty', 'like', 'go', 'get', 'let', 'play', 'many', 'put'. Figure 4.2 illustrates the change from before to after removing stopwords from positively- and negatively-rating reviews. Table 4 shows descriptive numbers for Call of Duty reviews comparing the situations before and after removing stopwords for both negatively and positively-rating reviews.

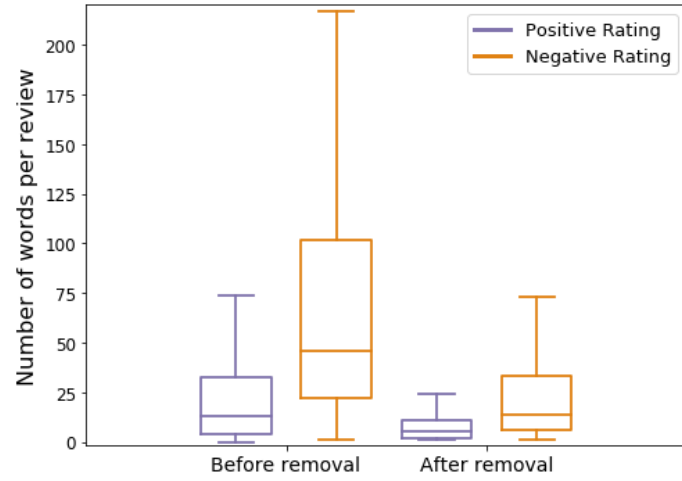
## 2. METHODOLOGY - LATENT DIRICHLET ALLOCATION

The data in this research focuses on one domain: video games. The data at hand is of textual nature. Instead of letting a human judge invest time in reading the data "by hand", the aim is to employ machine learning in order to find patterns in the data. These patterns may be recognizable in the data but they are abstract, thus calling for a method that is appropriate for uncovering latent patterns in data. One way to do this is by clustering. Clustering methods form groups of similar observations depending on their characteristics. They are unsupervised methods, that is, there is no prior knowledge of the true value of the output variable. Within clustering methods, a distinction is made between whether an observation can only belong to one cluster, or whether it can be part of multiple clusters simultaneously. The former would be a hard-clustering method, while the latter would be referred to as a soft-clustering method. In the case of the data at hand, it is of interest to generate an outcome that allows for observations (reviews, in this case) to belong to more than one cluster. This is because it is known that all the data is from the same domain (video games and even more specifically *Call of Duty: Ghosts*). The underlying goal of this section is to carve out what these customers writing these reviews care about. It is thus preferable to find relevant patterns that allow for more than a singular "meaning", i.e. cluster per review. Also, due to the nature of the data, it can be assumed, that multiple parts relevant to the domain may be referred to in one and the same observation.

One such soft-clustering method is Latent Dirichlet Allocation (LDA). Its aim is to reveal latent patterns in data. Applied to textual data, the distinguishable parts of these patterns are referred to as topics. LDA can be seen as a topic model. Intuitively, a topic encompasses elements that have an associative connection to one another. In the case of LDA, a topic is a latent variable uncovered by the model (Blei, Ng & Jordan, 2003) identified by different elements being prevalent per topic. LDA is a generative probabilistic method: Generative, because it generates an outcome, and probabilistic because it makes use of probabilities to arrive at the outcome. In the following the terminology of LDA will be briefly laid out. With textual data, each observation is referred to as a document, that is a collection of words. The collection of  $M$  documents is referred to as the corpus. Instead of referring to the elements contained in a document as words, they are referred to as terms:



**Figure 4.1:** Barchart showing the prevalence of star ratings among video game category reviews as compared to *Call of Duty: Ghosts* reviews



**Figure 4.2:** Boxplots illustrating summary statistics for the number of words per review before and after removing stopwords for positively- and negatively-rating reviews for *Call of Duty: Ghosts*

**Table 4:** Descriptive statistics for review data before and after removing stopwords.

	Positively-rating reviews			Negatively-rating reviews		
	Before	After	Change in %	Before	After	Change in %
Number of reviews	599	504	-15.9	374	364	-2.67
Sum of all words	21302	6541	-69.29	34705	10555	-69.59
Number of words per review:						
Maximum	969	310	-68.01	1326	400	-69.83
Minimum	1	1	0	1	1	0
Average	36	13	-63.89	93	29	-68.82
Median	13	5	-61.54	46	14	-69.57

a document consists of  $N$  terms. LDA operates on the assumption that a number of overarching topics is present in these documents characterized by prevalence of certain terms. LDA uncovers latent topic distributions from of an unstructured corpus onto a low-dimensional subspace, thus also shares capabilities of a dimension reduction methods. It follows a list of definitions of relevant terminology:

- $K$  is the number of topics.
- $\delta$  governs the per-topic term distribution  $\beta_k$ .
- $\alpha$  is a hyperparameter controlling the per-document topic distribution  $\theta_n$ .

The number of topics  $K$  has to be defined before running the model. The hyperparameter  $\alpha$  and  $\beta$  can be chosen to have a value between 0 and 1.  $\alpha$  controls the sparsity. That is, a high value for  $\alpha$  implies that every document contains a mixture of most topics available. As a corollary, a low value for  $\alpha$  would point towards documents that belong to few topics. A low value for  $\alpha$  could lead to overfitting. Similarly, a high value for  $\beta$  would imply that each topic likely contains a mixture of most terms, whereas a low value for  $\beta$  would imply that each topic only contains few of the terms. For topic  $k$ ,  $\beta_k$  is a vector of probabilities, each element representing the probability of a term to belong to that topic.

Intuitively, LDA assumes that the terms united in a document are related to one another. It tries to model the corpus to explain which terms appear together often (making up a topic). From that, it can infer what topics make up a document. LDA is a bag-of-words method, that is, it ignores the syntax and thus grammar of terms within a document. The output of LDA is a topic x term matrix. Row-wise, the elements sum to 1, since each row represents the probabilities of terms belonging to a topic. As such, a term can occur in multiple topics. As a corollary, a document can include multiple topics. From the nature of the output follows a strength of LDA: its interpretability. Topics can be visualized by plotting so-called wordclouds, that contain those terms with the highest probability of belonging to that topic. The sizes of the terms when displayed can be altered according to their relative probability, to enhance the view of which terms were most likely to appear in a topic.

For LDA as an unsupervised method, there is no ground truth in terms of labeled dependent variable that the model's outcome can be compared to. However, LDA can be evaluated by measuring topic coherence, a score that, "corresponds well with human coherence judgements" (Mimno, Wallach, Talley, Leenders, & McCallum, 2011). They intend for it to be used on domain-specific collections of documents. Several of the problems experts find with domain specific topics can be related to the concept of term co-occurrence. In relation to co-occurrence, the authors state that topics can be evaluated as of bad quality when: not all term chains (pairs of two terms) are sensible, intruder terms are detected within a topic or when the list of terms is not connected and random. The coherence score is calculated using the number of document frequencies ( $D(v)$ ) and the number of co-document frequencies ( $D(v, v')$ ) of different terms. It does so using a list  $V$  of the most probable terms  $T$  for a topic  $k$  such that  $V^{(k)} = (v_1^{(k)}, \dots, v_T^{(k)})$ . Topic coherence is defined as:



$$C(k; V^{(k)}) = \sum_{t=2}^T \sum_{l=1}^{t-1} \log \frac{D(v_t^{(k)}, v_l^{(k)})+1}{D(v_l^{(k)})},$$

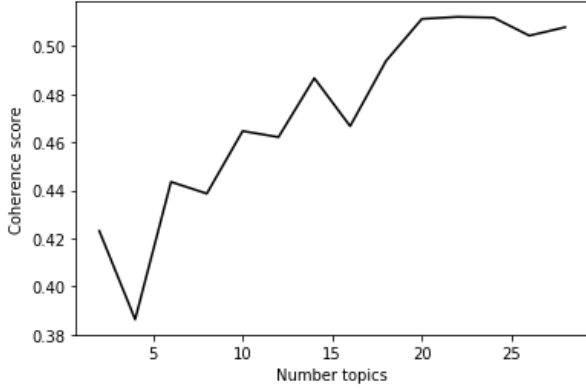
where 1 is added to account for the possibility of having a co-document frequency of 0 i.e. no co-occurrence of two terms in any document for that topic  $k$ . This would lead to taking the logarithm of 0, which is undefined. The coherence score for a topic lies between 0 and 1, where 1 indicates perfect coherence. Mimno et al. also note that the extensive use of co-occurrence information eliminates the requirement of a held-out corpus.

### 3. RESULTS

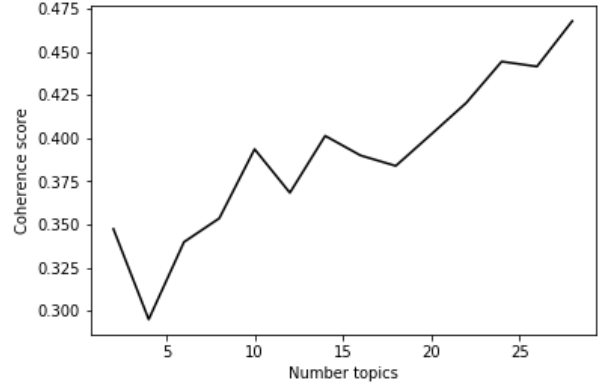
#### 3.1 Choosing the Models

LDA is first applied to the entire corpus of video game reviews for *Call of Duty: Ghosts*. An important input into the LDA model is  $K$ , the number of topics. It can be determined by using the coherence measure. LDA is run on a range of different values for the number of topics  $K$ . For each thus resulting model, the coherence measure is computed. The model with the highest coherence measure is chosen, and the  $K$  determined respectively.  $\alpha$  the hyperparameter controlling the per-document topic distribution is set as  $5/K$  as a rule of thumb.  $\beta$  was found through trial and error to be optimal at 0.01 and is set to that. In this section, LDA is applied to three different sets of documents. For each, coherence measure is used as the evaluation metric. First, LDA is applied to the entirety of the reviews for *Call of Duty: Ghosts*. However, it is of interest to see, which features of the game are liked and which are subject to criticism, after all one can assume that there is some consensus on liked versus disliked features of this specific game. In order to find these underlying characteristics, positively-rating reviews are examined separately from negatively-rating reviews. As such, LDA is secondly applied to all *Call of Duty: Ghosts* reviews with positive ratings. As a corollary it is last applied to all *Call of Duty: Ghosts* reviews with negative ratings.

When applied to the entire corpus of reviews (both positively- and negatively-rating reviews), LDA reaches the highest coherence score when generating  $K=8$  topics. The coherence score here is 0.4052 as seen in Table 5. This is the baseline model. Table 5 shows that the highest coherence scores for both other models, either using only positively- or negatively rating reviews are higher than that for LDA on all reviews. The evolutions of the coherence scores for both subsets of the reviews are displayed in Figure 4.3 and 4.4. For LDA on all positively-rating reviews (as seen in Figure 4.3), the highest coherence score attained is 0.51 for  $K=20$  topics. Considering that the coherence score falls between 0 and 1, this constitutes a 10% increase from the  $K=8$  baseline model. For LDA on all reviews with negative ratings, the highest coherence score attained is 0.44, as seen in Figure 4.4, leading to a choice of  $K=24$  topics. Though more slight than the increase from the baseline to the positively-rating reviews model, running LDA on negatively-rating reviews almost constitutes a 3% increase. Thus, examining the two subsets of reviews for *Call of Duty:*



**Figure 4.3:** Plot showing for reviews with positive ratings the coherence score evolution with respect to  $K$  the number of LDA topics



**Figure 4.4:** Plot showing for reviews with negative ratings the coherence score evolution with respect to  $K$  the number of LDA topics

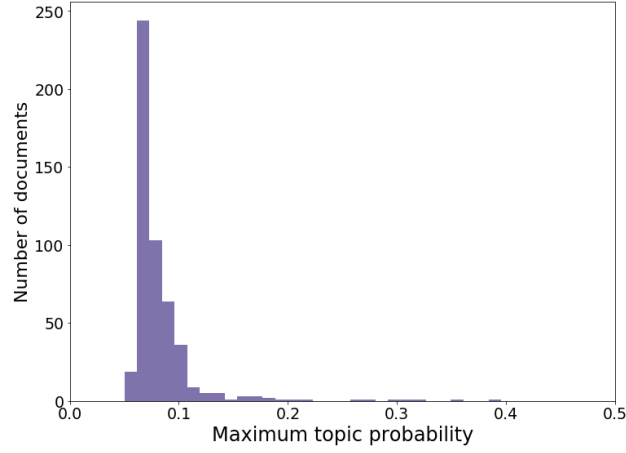
**Table 5:** Parameters and model performance for LDA on Call of Duty: Ghosts reviews.

	$\alpha$	$\beta$	Number of topics $K$	Coherence score
All reviews	0.10	0.01	8	0.41
Positively-rating reviews	0.25	0.01	20	0.51
Negatively-rating reviews	0.21	0.01	24	0.44

*Ghosts* results in overall higher coherence scores than looking at the entirety of the reviews at once. Considering this, the remainder of this section analyzes the topics respectively generated by LDA on positively- versus negatively-rating reviews.

### 3.2 Topics Inferred from Customer Reviews with Positive Ratings

In the following, the topics inferred by LDA on reviews with positive ratings are described, but only those that the researcher deems expressive. The non-conclusive ones will be skipped. Since for the LDA analysis on positively-rating reviews, the number of topics was chosen to be  $K=20$ , Figure 4.6 shows all 20 topic wordclouds and respectively, their ten top terms. Before going into detail on the possible meaning of certain topics, overall patterns in the topics are described. Figure 4.5 displays the maximum topic likelihood for all documents in the corpus of positively-rating reviews. That is, it shows how distinct the topic belongings elicited by this model are. As can be seen, the maximum topic likelihood in this case is densely clustered around just below 10%, with a sharp decrease after. It shows, that only a small number of documents feature a maximum topic belonging of over 10%. Thus, in this model, topic belongings are not very distinct and most documents contain a mixture of a number of topics. In Table 6 it can be seen that the topics with the smallest maximum contributions to a document lie around 10% while the highest maximum contributions are about 25% and up to almost 40%. Looking at the number of documents for which a topic is the most dominant among all topics, Table 6 shows that the highest counts are at above 30 documents. It is striking that the topics with relatively higher maximum contributions tend to be those with lower



**Figure 4.5:** Histogram showing for reviews with positive ratings the maximum topic likelihood for documents for  $K=20$

**Table 6:** LDA topics on positively-rating *Call of Duty: Ghosts* reviews

Topic	Maximum contribution to a document (in %)	Most dominant topic (count in documents)		
		4 stars	5 stars	Total
1	15.71	<b>10</b>	<b>34</b>	<b>44</b>
2	<i>10.01</i>	8	24	32
3	12.10	<i>3</i>	<b>30</b>	<b>33</b>
4	<i>10.54</i>	<i>1</i>	12	<i>13</i>
5	19.95	8	11	19
6	13.60	<b>10</b>	<b>48</b>	<b>58</b>
7	<b>31.28</b>	<i>4</i>	<i>6</i>	<i>10</i>
8	<b>35.18</b>	6	<i>5</i>	<i>11</i>
9	<i>10.29</i>	<b>17</b>	14	31
10	<b>39.54</b>	7	<i>8</i>	<i>15</i>
11	27.06	8	15	23
12	11.79	7	<b>39</b>	<b>46</b>
13	14.37	<b>14</b>	16	30
14	17.54	9	21	30
15	20.41	7	<i>8</i>	<i>15</i>
16	<b>32.11</b>	8	9	17
17	16.53	6	9	<i>15</i>
18	21.98	<i>5</i>	10	<i>15</i>
19	14.15	7	16	23
20	<i>10.72</i>	9	15	24

**Bold** signify the highest entries per column.

*Italics* signify the lowest entries per column.

dominant topic document counts. On the other hand, the topics with relatively lower maximum topic contributions seem to be those with higher dominant topic counts. This could imply that there are some topics which cluster documents about distinct topics, whereas there are other, more general topics that include a higher number of documents but are less distinct. Italicized words in the flowing text signify terms that are featured on the respective word cloud in Figure 4.6.

Overall, the topics elicited in this model seem to have four general themes: multiplayer mode, singleplayer mode as well as shipping and service. Topics 2, 7, 10, 11 and 13 resemble the multiplayer mode which is played online in Call of Duty games. Topic 2 has low maximum topic contribution and is the most dominant topic in a medium amount of documents. Looking at Figure 4.6, it seems to be about the *extinction mode* (represented by the bigram *extinction\_mode*), along with terms like *cool*, *amazing*, *awesome* and *cheap*, possibly indicating documents conveying a general sense of praise. Topics 7 and 10 both have high maximum topic contributions along with low dominant document counts overall. This suggests there is a specific group of documents that have high contributions from these topics. Topic 7 appears to be connected to the *large* size of the *map* and *campers* present in online mode. Campers are players who let their soldiers hide out waiting to kill another player who walks past. Topic 10 appears to be about technical enhancements the player can *purchase* for their figure. This is exemplified by the presence of *attachment* and *weapon* as well as *perk*. Perks are special skills given to the *player*. With leveling up comes more experience, and new *guns* and skills can be *unlocked*. Here, technical enhancements are improvements of the performance of the player, rather than e.g. speed of loading the game at large. With average contribution percentages and document counts, topic 11 looks to be about *zombies* that can be encountered in online *mode*. The presence of *challenge*, *worth* and *decent* could indicate appreciation. Topic 11 is relatively more prevalent among reviews giving the game 5 stars. Topic 13 is relatively present among reviews giving 4 stars. It features *multiplayer*, *improvement* and *dog*. This suggests that players could like the addition of these parts to the game, and could be said to give a rather general impression of the game.

The contents of topics 8 and 9 are likely to be about *singleplayer* mode. Topic 8 has high maximum topic contribution and very low most dominant document counts, suggesting it is uncommon but very prevalent in few documents. The overarching theme in this topic seems to be the campaign mode, where the player gets a *mission* along with a *storyline* to play by. Moreover, the presence of *improve*, *addition* as well as *launch* and *version* could point towards a positive view towards this installment of Call of Duty expressed through this topic. Topic 9 has low maximum topic contribution and is the most dominant topic in relatively many documents, especially 4 stars reviews. It likely refers to the length of the *story* in *singleplayer* mode in this part of the *series*, albeit *short*. Players are possibly *huge fans* of it and *feel* they can *recommend* it because it is *interesting*.

**Figure 4.6:** Wordclouds for 20 topics generated with LDA on Amazon customer reviews with positive ratings for the video game *Call of Duty: Ghosts*



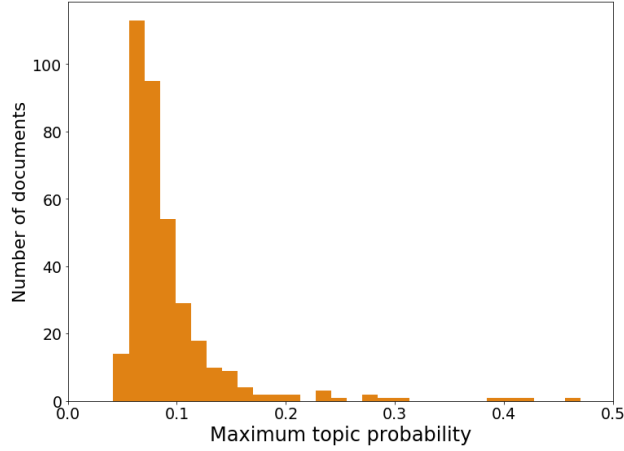
Another group of topics seems to be related to shipping, service and price of the game, namely, topics 1, 6 and 12. All three of them have maximum topic likelihoods between 10% and 16%, which is mid-range. Topics 1 and 6 are prevalent in both 4 and 5 star giving reviews, while topic 12 is the most dominant topic in 39 of reviews giving 5 stars, making it the second most prevalent. In topic 1, the prevalence of the terms *shipping*, *high(ly)* and *recommend* suggest satisfaction with Amazon’s shipping. *Lag* and *long* could point towards dispersed problems with the service, but this is non-conclusive. Topic 6 features *quality*, *price* and *fast\_shipping*. The terms *excellent*, *satisfied* and *super* suggest appreciation in reviews featuring this topic. Topic 12, almost exclusively dominating 5 star reviews, implies views on service as well: *condition*, *receive* and *gift*. *Love* and *awesome* could signify positive emotions. The terms *squad* and *storyline* seem intrusive and misplaced in this topic.

The most dominant word in topic 18 is *ghost*, a reference to the title of the game in question. It is relatively more prevalent in five star reviews than in four star reviews. The presence of *favorite*, *worth* and *variety* could imply an appreciation of not only the installment at large but also *sound* and *style* of the game. Overall *Call of Duty: Ghosts* reviews on Amazon with positive ratings (4 or 5 stars out of 5) seem to focus on the singleplayer (campaign) mode as well as multiplayer mode and are thus largely about gameplay, i.e. how a player perceives the game. This suggests that players appreciated at least parts of these modes enough to include them in their appraisal of the game. With respect to the product description, this creates the impression that the quality of the campaign as well as the perks gamers can add to their character in the game are relevant. Issues regarding how well the game works from a technical perspective seem to play a smaller role in reviews awarding *Call of Duty: Ghosts* positive ratings. Lastly, Amazon seemingly manages shipping well, as this was a recurrent term in the topics.

### 3.3 Topics Inferred from Customer Reviews with Negative Ratings

In the following, the topics generated using LDA on negatively-rating reviews are described, however, only if they are expressive. Figure 4.8 displays 24 wordclouds of the ten top terms on negatively-rating reviews. Before going into detail on the possible meaning of certain topics, overall patterns in the topics are described. Figure 4.7 shows the maximum topic likelihood for all documents in the corpus of negatively-rating reviews. It displays to what extent the documents feature distinct topic belongings. A peak at around 0.09 maximum topic probability implies that most reviews in question have rather mixed topic belongings, and are not dominated by one topic. That is, it shows how distinct the topic belongings elicited by this model are. Here, as with the model on positively-rating reviews, topic belongings are not very distinct but there is a less steep drop as maximum topic probability increases.

Table 7 displays that lowest maximum percentage contributions of the topics to a document lie below 14% while the highest maximum contributions range from 30% and up to almost 47%. Looking at the number of documents for which a topic is the most dominant among all topics, Table 7 shows that the highest counts are at above 30 documents. As in the previous subsection, an



**Figure 4.7:** Histogram showing for reviews with negative ratings the maximum topic likelihood for documents for  $K=24$

**Table 7:** LDA topics on negatively-rating *Call of Duty: Ghosts* reviews.

Topic	Maximum contribution to a document (in %)	Most dominant topic (count in documents)			
		1 stars	2 stars	3 stars	Total
1	<b>30.07</b>	3	<i>1</i>	8	12
2	<i>13.69</i>	4	<b>7</b>	7	18
3	<i>12.92</i>	4	4	<b>16</b>	<b>24</b>
4	29.75	<i>2</i>	4	4	10
5	20.42	4	4	<i>2</i>	10
6	<b>41.33</b>	8	3	4	15
7	17.83	<b>10</b>	5	3	18
8	<i>12.21</i>	4	<i>1</i>	6	11
9	14.02	<i>2</i>	4	5	11
10	27.10	3	<i>2</i>	4	<i>7</i>
11	14.42	<b>18</b>	<b>6</b>	4	<b>28</b>
12	14.71	3	7	<b>12</b>	22
13	<b>46.99</b>	<i>1</i>	<i>1</i>	4	<i>6</i>
14	17.24	5	3	<b>22</b>	<b>30</b>
15	<i>10.94</i>	4	4	6	14
16	<i>12.38</i>	<b>15</b>	<b>7</b>	5	<b>27</b>
17	16.86	5	4	6	15
18	22.82	4	<i>2</i>	9	15
19	<b>38.90</b>	<i>0</i>	4	<i>0</i>	<i>4</i>
20	20.09	6	4	5	15
21	27.06	9	<i>2</i>	<i>1</i>	12
22	14.47	<b>10</b>	<i>2</i>	9	21
23	24.38	7	<b>6</b>	<i>1</i>	14
24	<b>39.97</b>	<i>0</i>	<i>2</i>	3	<i>5</i>

**Bold** signify the highest entries per column.

*Italics* signify the lowest entries per column.

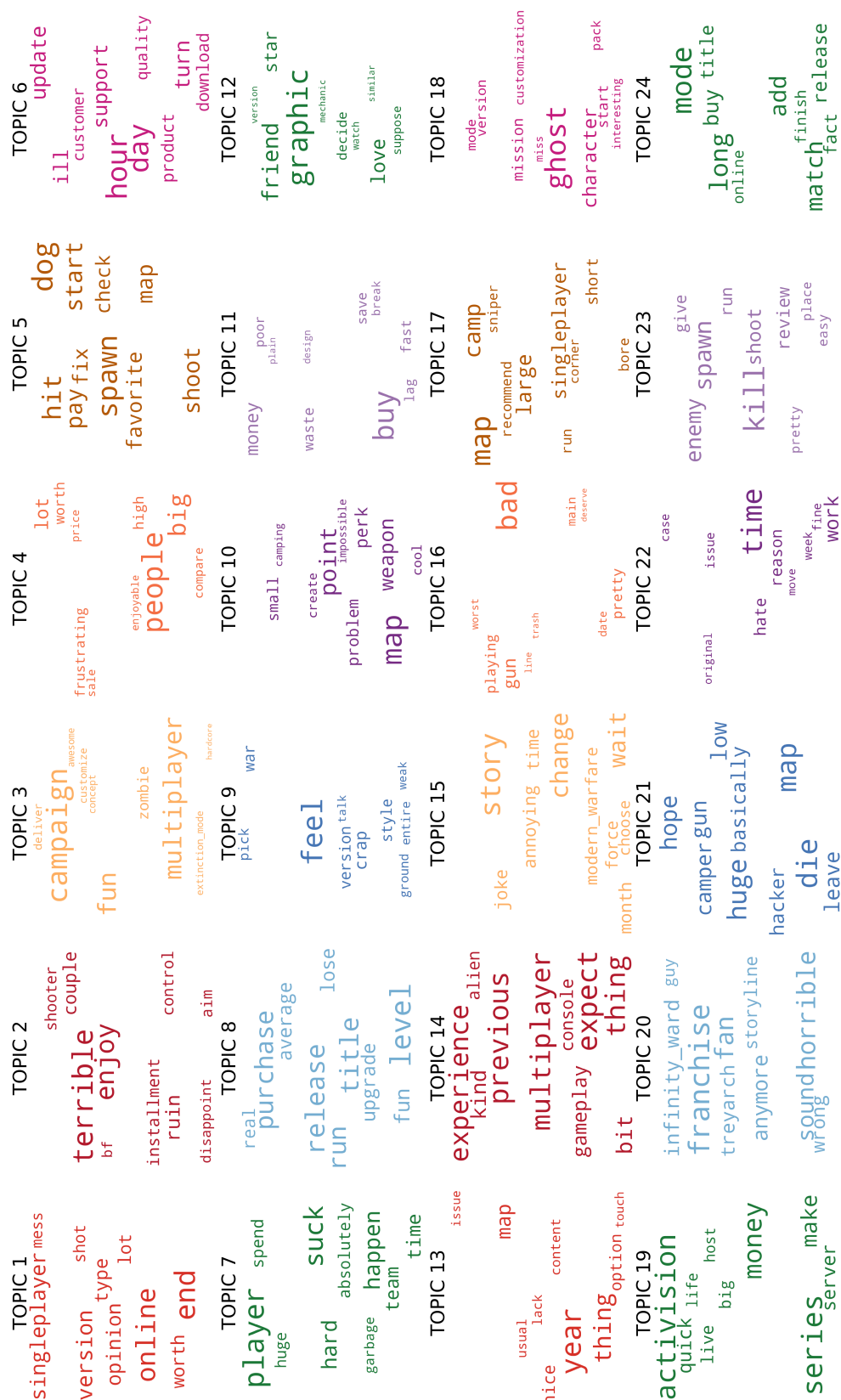
observation is that those topics with relatively higher maximum contributions tend to have lower dominant document counts. On the other hand, the topics with relatively lower maximum topic contributions seem to be those with higher dominant topic counts. This could imply that there are some topics which cluster documents about distinct topics, whereas there are other, more general topics that include a higher number of documents but are less distinct. Italicized words in the flowing text signify terms that are featured on the respective word cloud in Figure 4.8.

By and large, topics elicited in the model on negatively-rating reviews seem to include feedback on the multiplayer and some on the singleplayer mode. However, a considerable number of topics focuses on technical support coming from the developers of the game, hinting towards problematic features that are under attack in reviews. Topics 7, 10, 11, 17 and 21 appear to be about the multiplayer game mode. Topics 6, 11, 19, 20 and 21 are likely to be about the aforementioned technical support. Topic 7 is most prevalent among 1 star reviews. It features strong negatively connotated words like *garbage*, and *suck*. With it, *team*, *time* and *spend* are featured. Thus, this topic could refer to the *team* functionality in multiplayer mode. Topic 17 is about multiplayer mode (the term *singleplayer* seems intrusive here, unconnected to the other top terms). It includes the terms *large*, *map*, *camp*, *sniper*, *corner* and *bore*. Thus, reviews that likely belong to this topic could criticize the too large size of the map which makes it boring because it takes long to find enemies. Furthermore, the presence of *camp* could players waiting behind walls or *corners* to eliminate passers-by. This suggests that there is a portion of reviewers who do not mind campers.

Similarly to topic 17, topic 10 includes *map*, *weapon*, *perk* and *camping*. *Impossible* and *problem* suggest further dislike of players who camp. Topic 10 is relatively specific due to high maximum topic contribution and few overall documents where it is the most dominant topic. Topics 5 and 23 include *spawn*. Players spawn i.e. appear (seemingly randomly) on a *map* and fight each other online. In this scenario, speed and reaction time of the game (i.e. reliable server speed) is of the essence, unless the player wants to *lose* against the *enemy*. Which *place* on the map is important to be able to orient before getting *hit*, *shot* and *killed* too *easy* by an enemy. Topic 18 appears to be related to both the singleplayer and the multiplayer mode in the current installment of the franchise as suggested by: *ghost*, *version* and *mode*. *Customization*, *character* and *mission* point towards irritation with these parts of the modes.



**Figure 4.8:** Wordclouds for 24 topics generated with LDA on Amazon customer reviews with negative ratings for the video game *Call of Duty: Ghosts*



The following describes topics 11 and 21, which are both about the specific technical challenges experienced by players in the multiplayer mode. Topic 11 is dominant in 1 and 2 star reviews. Topic 21 is dominant in 1 star reviews. Together, they are the most dominant topics in over a fifth of 1 star reviews, suggesting technical challenges are a pressing issue. Topic 11 features the following: *money*, *waste* and *break* suggesting reviewers might describe the game as a waste of money. Further, *lag*, *poor* and *design* might refer to the quality of server speed and the deteriorated playing experience related to slow loading times and poor game design on the side of the developers. Topic 21 includes *hacker*, *map*, *die*, *camper* and *huge*. Again, reviewers possibly voice their concerns about camping, and furthermore, the presence of *hacker* points towards problems with hackers disturbing the gameplay. The following topics look to be related to technical issues as well as the developing company: 6, 19 and 20. Topic 6 has highly concentrated maximum topic likelihood. Presence of *hour*, *update*, *day*, *download*, *customer*, *support*, *ill*, *quality* and *product* clearly point towards reviewers being unhappy with *customer* service and *support* on the side of the game developing company. Topics 19 and 20 include on the one hand, the developers of the Ghosts installment in question, *activision* (which owns *infinity\_ward*) as well as the alternating game developer, *treyarch*. Topic 19 has high maximum topic likelihood and is only present in four 2 star game reviews. It includes *server*, *live*, *life*, *quick*, *host* and *activision* - this could indicate that reviewers complain about short soldier lives, due to slow servers operated by Activision hindering fast quick reaction times. Topic 20 includes general terms (*storyline*, *sound*) but seems to largely compare *infinity\_ward* with *treyarch* with respect to the Call of Duty *franchise*. Further included terms could suggest that the *horrible*, *wrong* performance of Activision makes the reviewers not a *fan anymore*.

Overall negatively-rating reviews seem to predominantly focus on technicalities regarding server running times and issues with the improvements and upgrades developers of the game manage to deploy. It appears that technical problems impact the multiplayer mode playing quality. There are also complaints regarding some players' strategy in the multiplayer mode, involving camping, which is a waiting strategy that is employed by some players, to surprise unsuspecting enemies. This might especially become a problem in large maps (also part of negatively-rating reviews) when players have to wait longer before they find enemies, than on a smaller map. Furthermore, it appears that the game development company Activision is compared to Treyarch, if not criticized even.

### 3.4 LDA Topics in the Product Description

For this subsection, the product description of *Call of Duty: Ghosts* on Amazon, found in Appendix ?? is divided into five parts, according to the section titles. For each section, first the LDA model from positively-rating reviews, and secondly, the LDA model from negatively-rating reviews is used to infer what topic of these respective model would contribute most to that section. As can be seen in Table 8, the contributions in % are all below 10%, except for LDA on negative reviews for section 4 Squads. When seeing what topics appear most dominant when applying the positive LDA model (refer to the visualization in Figure 4.6) on the parts of the description the following becomes

apparent:

- Section 1: Topic 13 is most dominant but a rather general one, interpreted as relating to the franchise at large.
- Section 2: Topic 14 is not considered interpretable. The model is not able to show what parts of the game mentioned in the description appear in positive reviews.
- Section 3: Topic 5 is unclear in terms of what is appreciated about spawning (the second most frequent term), though spawning is important in the multiplayer mode.
- Section 4: Topic 12 is considered interpretable, however, it is taken as rather related to the shipping process, than squads. The term squads appears seemingly at random, disconnected from the other terms.
- Section 5: Topic 13 is interpretable, though general. There is no visible connection with the content of this section.

The abovementioned shows, that the topics elicited by LDA do not conclusively point towards what parts of the description exactly are appreciated about the game. In the following, the most dominant topics per section from the negative LDA model (see Figure 4.8) are assessed:

- Section 1: Topic 20 is considered interpretable and is likely about the game developer companies that produce Call of Duty franchise, i.e. likely features comparisons between the two. Thus, this topic assignment can be considered appropriate, since this is the opening paragraph.
- Section 2 & Section 3: Topic 18 is considered interpretable and could refer to both single multiplayer mode. The presence of *customization*, *character* and *mission* points towards what themes are frequently found in reviews regarding single- and multiplayer.
- Section 4: Topic 24 is not considered interpretable and thus gives little to no information about what parts of the Squads in this installment of the franchise are disliked.
- Section 5: Topic 6 is considered interpretable and appears to comment on technical support. Though the content-related features of Extinction mode are clearly not referred to in this topic, the claim in the description to have, "fast-paced survival gameplay" seems risky; the topic most likely refers to parts of the customer service relating to the online mode that did not run smoothly.

### 3.5 So what? Implications for Description Formulation and Product Development

There are apparent differences between positively and negatively-rating reviews in terms of topics generated by LDA. The content of positively-rating reviews appears to be mainly related to well-liked storyline content in the campaign (singleplayer) mode, as well as parts of the online multiplayer (especially `extinction_mode`), such as weapons and perks. Moreover, these reviewers

**Table 8:** Table showing what topics had the maximum likelihood per section from the LDA models on positively and negatively-rating reviews.

Section title	LDA on positive reviews		LDA on negative reviews	
	Dominant topic	Contribution (in%)	Dominant topic	Contribution (in%)
1 Outnumbered	13	9.90	20	8.17
2 Singleplayer	14	8.14	18	8.59
3 Multiplayer	5	9.24	18	8.81
4 Squads	12	9.22	24	10.78
5 Extinction	13	7.56	7	7.18

seem to appreciate the service surrounding shipping of the game. Negatively-rating reviews, on the other hand, strike as critical of a strategy of players related to big maps (camping) as well as the short length of the campaign storyline in singleplayer mode. These reviews seem to insinuate dissatisfaction with the spawning system, i.e. it being flawed in such a way that players landing on a map were often killed by other players camping out. The term *spawn* is also featured in one topic in Figure 4.6, but in a non-interpretable topic. This suggests that while there might be some players appreciating the spawning system, concern about it is formulated more clearly. But most importantly, negatively-rating reviews appear to state disapproval of the quality server performance and game developers' (Activision) abilities to implement upgrades. The former is problematic in multiplayer mode, because it is played online and lags in server speed can lead to being killed faster, thus making players lose the game. The latter also applies to multiplayer mode, because game developers are supposed to upload updates when there are issues in the game. It is possible that players are thus criticizing the developers of moving the game to market too quickly, or having focused on the wrong details. This following paragraph is concerned with answering the hypothesis

H2.1: Differences in content between positively and negatively-rating customer reviews can inform a rephrasing of product descriptions.

What do the LDA findings regarding positively and negatively-rating reviews imply? This is with respect to the product description of *Call of Duty: Ghosts* on Amazon, found in Appendix ???. Overall, it seems that at least some referral to the content from most of the five section subheadings in the description are present in interpretable topics generated by LDA. The Single Player Campaign subsection speaks about the story to be experienced by the players. LDA does not appear to yield much information about the specific elements of the story, but seems to give the insight that the length of the storyline was an issue worthy enough for players to include it in their reviews. However, length is not mentioned in the description, therefore, it might be advisable to include an estimation of length in hours in the description.

The Multiplayer subsection of the product description specifically mentions the extent of customization possible within the game. The word customization occurs in negative LDA topics, but it is not apparent what part thereof players do not appreciate, i.e. the meaning is unclear. The positively-rating reviews do circle around the add-on features made available for weapons during the game, such as perks. Interestingly, the product description goes into much detail about the possibility of changing the gender of the soldier and their physical appearance, something that is entirely overlooked in the topics found by LDA and thus potentially not a selling point to be prioritized to the current customer segment. Being able to change the gender of the soldier, is a first for any version of the Call of Duty franchise, yet lacking in the customer reviews. This could point towards the fact that present customers do not care about this feature. The dog Riley, mentioned in the product description, does not appear to be a feature of the game that reviewers have a strong opinion on. It is striking that terms mentioned in the product description as specific to the Ghosts installment of the Call of Duty franchise, such as "ODIN" or "Juggernaut Maniac", are nowhere to be found in either corpus' LDA topics. This could mean either, that LDA was unable to uncover any latent patterns relating to these game specifics, or, that reviewers are at least indifferent about these terms and are more in it for the game of weapons and avoiding to get killed, than such specific labels and story details.

Phrases mentioned the subsection Extinction, appear several times in LDA topics on the two corpora. It is the extinction\_mode bigram, and the related zombies, that are the focus of this section in the description, so this is an overlap between reviews and the product description. Looking at the paragraph - and knowing that one of the biggest weaknesses reviewers complained about regarding *Call of Duty: Ghosts* refers to server performance - the promising mention of, "fast-paced" might cause great frustration with gamers. This is a clear example for a part of the product description that should be altered in response to examining the review feedback, lest the credibility of the entire description or even Activision might suffer. The same holds for the mention of, "60 frames-per-second across all platforms" in the first section of the description. This is a promise that the developers of the game apparently were unable to hold.

The content from subsection Squads, is largely missing among the interpretable LDA topics. The word squad does appear as a stray term in one LDA topic on positively-rating reviews; however, it is unclear whether this relates to camping teams or those referred to in the description. The term teams appears in a negative LDA topic along with several negatively connotated terms (*garbage*, *suck*). It is worth noting, that technical features relating to the performance of the game are, by and large, ignored in the description. Graphic and sound of the game seem to split opinions, since they both appear once, respectively in the topics of the two corpora. It can therefore not be deduced, whether users appreciate these in the Ghosts installment, or not. However, *sound* appears together with *ghosts* in positive LDA topic 18, and together with the names of the game developing companies in the negative LDA topic 20. This implies that the game sound might be compared from one installment of the franchise to another, and thus one developing company to

another. Therefore, sound could be a feature that is used to judge the quality of the installment by and large, making it central to the performance of the game. Thus, if special care went into improving sound, it should be mentioned in the description to inform repeat buyers of the franchise. Overall, consulting the Amazon product description and comparing it on LDA from the two corpora, shows that there are clear incongruences between what feedback reviewers talk about and what the description mentions, stresses and praises. It can be concluded that there are opportunities for adjustments to the description based on separating positive from negative reviews, and that the LDA topics are conclusive enough to accept H2.1. The following is concerned with answering

H2.2: There is underlying information in customer reviews that points towards keeping, removing or enhancing certain aspects of a product for future versions.

So, what does this mean with respect to product development? Since in the Call of Duty franchise, Activision and Treyarch take turns in developing and publishing a game, Treyarch could benefit from the criticisms of the *Call of Duty: Ghosts* game in the long run as long as it manages to at least outperform Activision and meets customer expectations. The single biggest issue seems to be server speed and reliability of post-publishing game update quality. Since players know that the next game will be developed by the other firm, Treyarch should invest time during pre-production in taking the negative performance of Activision's game and turning it around to show the players they are the superior developer. Moreover, singleplayer mode is criticized for being too short. In order to cater to all players, also those who prefer the singleplayer containing the campaign mission and storyline, it could prove to be wise to invest in a longer storyline. At the same time, Treyarch should pay close attention to the features players that reviewers directly mentioned together with the game developers company names and the Ghosts installment specifically. That is, the sound and the content of the storyline. Sound in this game contributes to the realistic nature and immersive quality; it cannot be fixed in post-production, thus must be on point at the time of publication. If a campaign story is long but not enjoyable, it could backfire on Treyarch, so the content of the storyline should be well thought through. In the advertising of their subsequent version of the franchise, Treyarch should then try to incorporate high server speed and long storyline as well as an improved spawning system to attract customers. As mentioned previously, LDA topics from the two corpora did not suggest an interest from reviewers regarding the newly-introduced option to decide the soldier's gender. While for Activision, this could be considered a "flop", Treyarch could use the fact that existing players are now used to the presence of that feature. In fact, they could keep the feature intact, while utilizing it to create an attractive proposition for new target groups, i.e. those that might be interested in playing a female character. Seeing as the existing customers do not care, the marketing message including the gender option should be different from the "main" marketing message.

As seen in subsection, 3.4 the findings from applying LDA on the section documents from the Amazon product description have three main implications. First, that details regarding story

content are not elicited from reviews. Thus, LDA does not reveal which elements of the storyline in campaign mode are either appreciated or disliked the most (depending on which topic model). Similar findings apply to the multiplayer mode; specific terms and customization features unique to this installment of *Call of Duty* are not elicited from LDA on reviews. Third, and perhaps most striking is the topic most dominant in the extinction section from the negative LDA model. The product description there mentions fast-paced gameplay, while the topic associated points towards slow updates and bad customer support. So the claim of speedy servers seems bold. This shows that it is possible to filter out relevant topics from a body of reviews using LDA to inform game development and marketing for a specific game. While this should not be the only deciding factor in research on customer satisfaction with a product, it has the potential to be a valuable addition. It is demonstrated, that descriptions can and should be readjusted after the initial sale to shine a light on the more positively perceived aspects of the game, while avoiding to reflect the more negatively-perceived parts. Furthermore, as was seen in the LDA on negative *Call of Duty: Ghosts* reviews, customer service was a relevant issue, something that could be tackled in post-production with more immediacy than publishing a new game. As a result of this, the researcher accepts H2.2.

Lastly, it can be said that at least partially, the topics elicited using LDA are similar to those elicited using the human judgement approach. This counts for storyline and modes of game, so especially the attributes, rather than the consequences. Furthermore, an analysis similar to the one in this research could be employed by competing developers or producing firms to enhance their own games and marketing communication based on the weaknesses in *Call of Duty: Ghosts*.

## 5. GENERAL DISCUSSION

The research question this paper aims to answer is as follows:

How can UGC be exploited to elevate the impact of marketing communications and inform product development?

Looking at the product category of video games, this research establishes the existence of an incongruence between product descriptions and customer reviews. It also exemplifies how video game production firms could alleviate this incongruence, using the individual example of the video game *Call of Duty: Ghosts*, chosen because it received a large body of mixed reviews. In the first part of this research, the incongruence between the product descriptions and customer reviews for the video game category at large as present on Amazon is established. This is executed using the means-end chain laddering approach. Instead of using costly interviews, this research made use of a sample of product descriptions. A sample of these in the relevant domain is examined by a human judge to elicit product attributes and consequences common and relevant for video games. Based on Kahle's List of Values, personal values common to individual humans, i.e. consumers, belonging to each of the consequences elicited are appointed. Subsequently, the patterns of appearance of

said attributes and consequences in descriptions is compared to their pattern of occurrence in the respective customer reviews. Their occurrence in customer reviews is found consulting human judges on Amazon MTurk. Matching present attributes and consequences for both forms of text is necessary in order to establish the incongruence, i.e. considerably incoherent occurrence between respective descriptions and reviews. Applying MCA, it is found, that there are considerable differences in the prevalence of attributes and consequences as well as the general pattern of occurrence.

This paragraph extrapolates the findings from MCA to the patterns exhibited by values and contrasting this between product descriptions and customer reviews. Values are connected to attributes and consequences through the means-end chain. Generally, the relevant dimensions of descriptions appear to contrast two concepts; on the one hand, function, action and doing and on the other hand, a sense of feeling, form and design. The former insinuates values related to respect and accomplishment. Conversely the latter is reflects values of belonging and self-fulfillment. Coming to the content of reviews, there is less contrast between the dimensions. Reviews create the impression that customers are mainly concerned with values of self-fulfillment versus belonging. Thus, the analysis of reviews suggests that customers are more concerned with voicing attributes and consequences related to the emotional and softer values, exemplified by the second dimension of MCA on descriptions. Following the preceding paragraphs, H1 is accepted. It follows, that sub-question 1 (Is there an incongruence between product descriptions and customer reviews?) can be confirmed.

The second part of this research is concerned with exploring the incongruence between message sent (product descriptions) and feedback given (customer reviews). It first does so by singling out a product from the video game domain: *Call of Duty: Ghosts*, in order to imitate the process a video game production might follow. Using LDA, one aim is to establish whether customer reviews with positive ratings of 4 or more stars reveal different latent topics than those with negative ratings of 3 or less stars and moreover, whether this information can be exploited to adapt the formulation of the product description. Such a rephrasing can be argued to have an impact on relevance and efficacy of the promotion of the game as well as the credibility production firm at large. It is found that in the case of *Call of Duty: Ghosts*, the producing firm Activision should defer attention away from the pace of the game (since reviewers experienced server issues) and toward sound quality, if there are significant improvements compared to the last installment of the franchise. Furthermore, several details in the description are not found in any of the LDA topics, for example, the mentions of version-specific terms "ODIN" and "Juggernaut Maniac". Either LDA was insufficient to discover all patterns, or, customers did not care about them enough to refer to them in their reviews. These thus possibly represent superfluous details. The findings from contrasting LDA on the two review corpora therefore suggest rephrasing of the product description, based on the assumption that gamers care about what they write in reviews. Therefore, H2.1 is accepted.

As a note to the phrasing of product descriptions, text length is a point worth considering. In line



with theory surrounding cognitive load, whether or not it is in the best interest of a producing company to trigger normative decision making depends in part on the performance of the product on the market. Drolet et al. (2004) found that higher cognitive load can mitigate trade-off avoidance, triggering normative decisions. For a well-received product (as shown by largely positive reviews), this suggests employing large texts, in order to not give readers i.e. potential customers opportunity to begin doubting their purchase decision. On the other hand, if a product visibly attracts more animosity, it might be advisable to create relatively short product descriptions, decreasing cognitive load. This could make room for developing trade-off avoidance. Since trade-off avoidance leads to less normative decision making, it might confuse the potential customer and blur the negative feedback visible on the page, possibly causing them to make a purchase. Thus, for a game that received relatively bad reviews like *Call of Duty: Ghosts*, a shortened product description strikes as being a good tactic.

Lastly, it is established that underlying information in customer reviews can be used to inform future product development. This has high relevance in the video game world, and especially the Call of Duty franchise. A new version of Call of Duty is released every year, by alternating publishers Activision and Treyarch. Each firm could exploit findings from reviews regarding their alternating company to not only rephrase marketing communication, but also, invest in areas the other has shown (and now exemplified) weaknesses in. This must be done bearing in mind that the risk involved of customer expectations are not met. Thus, strengths of the preceding game should at least be matched while its weaknesses should be outperformed. Server speed should be invested in heavily, surpassing the flawed quality thereof in *Call of Duty: Ghosts*. Additionally, Activision advertised the introduction of a gender-changing feature. This appeared to be overlooked by reviewers. This has two possible implications. First, that current customers do not consider it important enough to mention. Second, that the alternate Call of Duty production firm Treyarch could exploit the fact that Activision introduced this feature for the following version developed by them. On the one hand, by creating a two-sided marketing strategy, one advertising to existing customers and the other hand, an adapted message to appeal to a new target group, one that would be interested to play as a female soldier. Based on the above findings H2.2 is accepted. Thus, a two-folded and affirmative answer is provided to sub-question 2 (How can customer reviews be used as grounds to adjust product descriptions and advise product development?).

## 6. LIMITATIONS AND FURTHER RESEARCH

Video games in particular lend themselves to the exploration of a plethora of topics from technicalities to more qualitative and creative features. A video game is similar to a book because it aims to capture the player/reader enough to immerse themselves in the story. Differently from a book, however, a video game is also judged on functionalities and features. Since different games have different settings, a description or a review may feature words that are related to the setting rather than the opinion expressed by the user or marketer. That is, a the story line of

a game might circle around a violent theme, such as warfare. Negatively connotated words used in combination with such a story line thus would not refer to the quality but to the nature of the game.

An assumption made when employing the means-end chain, was to treat product descriptions like interviews when eliciting attributes and consequences from descriptions. This was done in order to gain a general view of the product category of video games. Not individual games were the aim in the eliciting process, but rather drawing up an overview of video games at large. In hindsight, certain attributes that were established, such as *single and multiplayer* could have been split up into two parts. Conversely, the removal of the consequence mature brings up questions regarding how granular the analysis of the category-wide incongruence was, and whether it is truly even more striking, since some nice product features could have been overlooked or removed. This is connected to the limitations in using a human judge, since they exhibit bias.

As indicated by a worker via email to the researcher, it might be advisable to avoid grouping of several reviews into one HIT. Requesting the classification of five reviews turned out to be tiring to the workers on MTurk. The researcher partially blames this on the fact that she was not aware of the fact that the same worker can claim and complete several tasks of the same HIT. Thus, the workers might have been more motivated to complete more, but shorter HITs at a lower pay than 0.40\$. To possibly decrease the threatening nature of the HIT, it could have been mentioned that all five reviews were on the same product. Another way to decrease the workload per HIT could have been to only include attributes. This would have also removed some of the misunderstandings regarding what exactly constitutes the presence of a consequence (some workers were accepting synonyms of stated consequences, others not).

Concerning the second part of analysis, exploring the incongruence, there are limitations. Firstly, the choice of stop words could possibly bias the results. Moreover, there are assumptions that were made that could be challenged. Namely asking, if a review was accompanied by a negative rating, then is it safe to assume that the entirety of its content attacks the video game? And that in return, a review given with a positive star rating defends and/or praises the topics in the review?

An interesting question in the age of information is whether any product has to still be experienced before knowing what it is. There is a plethora of available videos on YouTube, influencer content on Instagram and free word-of-mouth across all channels of social media. The focus on product reviews alone is almost atomic, however, a good start to look more closely into the topic of the creative consumer. It would also be interesting to contrast the approach taken in this research regarding hedonic versus utilitarian goods. Lastly, implementing the improvements mentioned in the above could improve further research. For example, the stability of the collected data in the means-end chain laddering approach could possibly be improved by gathering a larger dataset or implementing clearer instructions for the survey.

If the approach of eliciting product features, i.e. attributes and consequences, using human judges could be automated, large studies of the current market structure could be done. In the long term, such automatization could save time and costs, as well as giving the opportunity of assessing developments over time. That is, it could give insights on current offerings, which parts of the market are saturated, what new (niche) trends can be observed and how prevalence of certain features develops over time. Moreover, an market structure analysis using reviews could give rise to the development of new products, using the patterns of what customers mention in reviews to see what features are closely linked in their minds. This would imply treating the customer as a prosumer.

In the interest of efficiency, it might be argued that the high number of topics generated using LDA negatively impacts said efficiency. Furthermore, the fact that different applications of LDA on different corpora result in different configurations and numbers of topics, could imply that the LDA is unstable. While this may be a valid point of critique another goal of this research was to filter out relevant opportunities for product development. Since these could be latent, a higher number of topics is not detrimental, but could be rather helpful. Furthermore, if there is reasonable amount of consensus among the reviewers of the positively-rating and the negatively-rating corpora, then it can be assumed that the two cohorts have separate fields of discussion: different topics of conversation and thus also possibly different number of topics. Additionally, a general theme such as server issues have manifold impacts on the game experience. It could be that some reviewers complain about themselves dying quickly, and relate that to the spawning system; it could be that some reviewers refer to the performance of the gaming developer, Activision and in turn talk about their developers and approach to updates. This illustrates that while there are overarching topics such as server issues or storyline, they may be interpreted in different ways with different applications and conclusions by different cohorts.

Whereas this research focuses on UGC in the form of Amazon customer reviews, a company could choose to use additional sources of customer feedback. For example, a list of all online shops selling the product of interest could be created and UGC from these pages could be used to extend the body of reviews. Moreover, social media channels like Twitter, Facebook or Instagram could be scraped for feedback content about the product. App stores could be used to analyze reviews for mobile games. YouTube has become a popular platform to share (sponsored) review videos for products, especially experience goods. Here, not only the review video itself, but also the comments could be gathered to create a more complete image of customer feedback on the product. As said before, this feedback can be used to update product descriptions but also as a way to help prioritize needed changes or opportunities for innovation.

Another opportunity for further research is examining the interplay between product description length, respective customer ratings and purchase decisions. This could be done by a survey. For it, ratings could be averaged per product. Then, the description would be displayed, accompanied by the average rating and the reaction of the survey taker could be recorded, i.e. purchase or not.

This could feature altering the length of a description for a product and presenting different lengths to different survey participants, thus manipulating the length and seeing, whether the simulated "purchase decisions" are affected by the length of the description. Similarly, the average rating could be altered, to examine whether it has an impact on the purchase decision. Such a study could be used to make inferences about the impact of cognitive load on trade-off avoidance and purchase decisions, and would be applicable to a plethora of product categories.

## 7. REFERENCES

- Abdi, H., & Valentin, D. (2007). Multiple correspondence analysis. *Encyclopedia of measurement and statistics*, 2, 651-66.
- Acher, M., Cleve, A., Perrouin, G., Heymans, P., Vanbeneden, C., Collet, P., & Lahire, P. (2012, January). On extracting feature models from product descriptions. In *Proceedings of the Sixth International Workshop on Variability Modeling of Software-Intensive Systems* (pp. 45-54).
- Amabile, T. M. (1983). The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology*, 45(2), 357.
- Amazon Reviews pds. (n.d.). Retrieved from <https://s3.amazonaws.com/amazon-reviews-pds/tsv/index.txt>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan), 993-1022
- Bragge, J., & Storgårds, J. (2007, September). Utilizing Text-Mining Tools to Enrich Traditional Literature Reviews. Case: Digital Games. In *Proceedings of the 30th Information Systems Research Seminar in Scandinavia* (pp. 1-24).
- Brody, S., & Elhadad, N. (2010, June). An unsupervised aspect-sentiment model for online reviews. In *Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics* (pp. 804-812). Association for Computational Linguistics.
- Brown, S. L., & Eisenhardt, K. M. (1995). Product development: Past research, present findings, and future directions. *Academy of Management Review*, 20(2), 343-378.
- Chandler, H. M. (2013). *The game production handbook*. Jones & Bartlett Publishers.
- Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*, 49(10), 1407-1424.
- Drolet, A., & Frances Luce, M. (2004). The rationalizing effects of cognitive load on emotion-based trade-off avoidance. *Journal of Consumer Research*, 31(1), 63-77.
- Fang, E. (2008). Customer participation and the trade-off between new product innovativeness and speed to market. *Journal of Marketing*, 72(4), 90-104.

- Griffin, A., & Hauser, J. R. (1993). The voice of the customer. *Marketing Science*, 12(1), 1-27.
- Hauser, J. R., & Clausing, D. (1988). The house of quality. *Harvard Business Review*.
- He, R., & McAuley, J. (2016, April). Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *proceedings of the 25th international conference on world wide web* (pp. 507-517). International World Wide Web Conferences Steering Committee.
- Hermes, I. (2018). *Attitudes towards concessions against unemployment*. Unpublished Manuscript. Erasmus University Rotterdam, Rotterdam, Zuid-Holland.
- Hirschman, E. C. (1980). Innovativeness, novelty seeking, and consumer creativity. *Journal of Consumer Research*, 7(3), 283-295.
- Hoffman, D. L., & De Leeuw, J. (1992). Interpreting multiple correspondence analysis as a multidimensional scaling method. *Marketing Letters*, 3(3), 259-272.
- Kahle, L.R., (1983), *Social Values and Social Change: Adaptation to Life in America*, New York: Praeger.
- Kahle, L. R., & Kennedy, P. (1988). Using the list of values (LOV) to understand consumers. *Journal of Services Marketing*, 2(4), 49-56.
- Katz, D. (1960). The functional approach to the study of attitudes. *Public Opinion Quarterly*, 24(2), 163-204.
- Lee, T. Y. (2007, August). Needs-based analysis of online customer reviews. In *Proceedings of the ninth international conference on Electronic commerce* (pp. 311-318)
- Lee, T. Y., & Bradlow, E. T. (2011). Automated marketing research using online customer reviews. *Journal of Marketing Research*, 48(5), 881-894.
- Lettl, C. (2007). User involvement competence for radical innovation. *Journal of Engineering and Technology Management*, 24(1-2), 53-75.
- Locander, W. B., & Spivey, W. A. (1978). A functional approach to attitude measurement. *Journal of Marketing Research*, 15(4), 576-587.

- McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. *Harvard Business Review*, 90(10), 60-68.
- McAuley, J., Targett, C., Shi, Q., & Van Den Hengel, A. (2015, August). Image-based recommendations on styles and substitutes. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 43-52). ACM.
- Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. (2011, July). Optimizing semantic coherence in topic models. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 262-272). Association for Computational Linguistics.
- Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy*, 78(2), 311-329.
- Nelson, P. (1974). Advertising as information. *Journal of Political Economy*, 82(4), 729-754.
- Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, 31(3), 521-543.
- Ramage, D., Dumais, S., & Liebling, D. (2010, May). Characterizing microblogs with topic models. In *Fourth international AAAI conference on weblogs and social media*.
- Ramsoy, T. Z. (2015). *Introduction to neuromarketing & consumer neuroscience*. Denmark: Neurons Inc Aps.
- Reynolds, T. J., & Gutman, J. (1988). Laddering theory, method, analysis, and interpretation. *Journal of Advertising Research*, 28(1), 11-31.
- Rindfleisch, A., & Moorman, C. (2001). The acquisition and utilization of information in new product alliances: A strength-of-ties perspective. *Journal of Marketing*, 65(2), 1-18.
- Schuff, D., & Mudambi, S. (2012). What makes a helpful online review? A study of customer reviews on Amazon. com. *Social Science Electronic Publishing*, 34(1), 185-200
- Spivey, W. A., Munson, J. M., & Locander, W. B. (1983). Improving the effectiveness of persuasive communications: Matching message with functional profile. *Journal of Business Research*, 11(2), 257-269.

Surowiecki, J. (2005). *The wisdom of crowds*. New York: Anchor Books.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257-285.

Timoshenko, A., & Hauser, J. R. (2019). Identifying customer needs from user-generated content. *Marketing Science*, 38(1), 1-20.

Toffler, A., & Alvin, T. (1980). *The third wave*. (Vol. 484). New York: Bantam books.

Vriens, M., & Hofstede, F. T. (2000). Linking Attribute, Benefits, and Consumer Values. *Marketing Research*, 12(3).

Welch, G. S. (1975). *Creativity and intelligence: A personality approach*. Chapel Hill, NC: Institute for Research in Social Science, University of North Carolina.



## A. APPENDIX: AMAZON MTURK SURVEY

Classify the content of video game reviews			
Requester: Eliza Reid	Reward: \$0.40 per task	Tasks available: 0	Duration: 1 Hours
Qualifications Required: Location is US , Number of HITs Approved greater than 1000 , HIT Approval Rate (%) for all Requesters' HITs greater than 95			

### Task

Determining the content of video game product reviews.

### Instructions

A **review text** can mention

1. Product attributes (PA): characteristic of product like price or color
2. Product consequences (PC): experienced as a result of using the product

Below is the **list** of product attributes and consequences (PA and PC) for **video games**.

#### Attributes

Modes of Play  
Single and Multiplayer  
Storyline  
Sound  
Characters  
Graphics  
Team Manager  
Customization

#### Consequences

Exciting  
Interactive  
Realistic  
Control over Game  
Save the World  
Prove Skills  
Identification  
Mature  
Responsibility  
Creative  
Nostalgia

After reading the **review text**, there are two steps:

1. Identify **matches** and check the boxes that apply.

A match is defined as a PA or PC that is mentioned in the **review text** *and also* present in the **list**.

2. Identify **mismatches** and write down in 1-3 words.

A mismatch is defined as a PA or PC that is mentioned in the **review text** *but not* present in the **list**.

#### EXAMPLE

- bold: **match**
- highlighted: **mismatch**

Fear Effect is excellent! The **graphics** are fun, the **storyline** is great, the gameplay is **exciting**... It is the best game I have played in a while. THE DOWNSIDE: The game has very little **replay value**, and it is not very **long**. It **only took a few hours to beat**. The game IS fun, though -- once I beat it, I went right back to play through in **Hard Mode**. After that, though, there was nothing more to the game. VERDICT: You should definitely rent this game if it interests you at all (you will not be disappointed). I would not recommend buying it, though.

**Mismatches:** replay value, length, difficulty

#### Instructions



[View full instructions](#)

##### Match

Product attributes and consequences that are in the review *as well as* the list: **check the boxes that apply**.

##### Mismatch

Product attributes and consequences that are in the review *but not* the list: **type in field**.

#### Review Text (1/5)

\$(reviewText1)

- |   |  |   |
|---|--|---|
| <input type="checkbox"/> Modes of Play          | <input type="checkbox"/> Exciting          | <input type="checkbox"/> Responsibility |
| <input type="checkbox"/> Single and Multiplayer | <input type="checkbox"/> Interactive       | <input type="checkbox"/> Creative       |
| <input type="checkbox"/> Storyline              | <input type="checkbox"/> Realistic         | <input type="checkbox"/> Nostalgia      |
| <input type="checkbox"/> Sound                  | <input type="checkbox"/> Control over Game |   |
| <input type="checkbox"/> Characters             | <input type="checkbox"/> Save the World    |   |
| <input type="checkbox"/> Graphics               | <input type="checkbox"/> Prove Skills      |   |
| <input type="checkbox"/> Team Manager           | <input type="checkbox"/> Identification    |   |
| <input type="checkbox"/> Customization          | <input type="checkbox"/> Mature            |   |

Mismatch

\_\_\_\_\_

## B. APPENDIX: PRODUCT DESCRIPTION OF CALL OF DUTY: GHOSTS

### **Outnumbered and Outgunned, but Not Outmatched**

Call of Duty: Ghosts is an extraordinary step forward for one of the largest entertainment franchises of all-time. This new chapter in the Call of Duty franchise features a new dynamic where players are on the side of a crippled nation fighting not for freedom, or liberty, but simply to survive. Fueling this all new Call of Duty experience, the franchise's new next-gen engine delivers stunning levels of immersion and performance, all while maintaining the speed and fluidity of 60 frames-per-second across all platforms.



#### **Single Player Campaign**

Ten years after a devastating mass event, America's borders and the balance of global power have changed forever. As what's left of the nation's Special Operations forces, a mysterious group known only as "Ghosts" leads the fight back against a newly emerged, technologically-superior global power. A New Call of Duty Universe: For the first time in franchise history, players will take on the underdog role with Call of Duty: Ghosts; outnumbered and outgunned, players must fight to reclaim a fallen nation in an intensely personal narrative. Gamers will get to know an entirely new cast of characters and visit locales in a changed world unlike anything seen in Call of Duty before.



#### **Multiplayer**

In Call of Duty: Ghosts you don't just create a class, you create a soldier, a first for the franchise. In the new Create-A-Soldier system, players can change the physical appearance of their soldier by

choosing the head, body type, head-gear and equipment, and for the first time in a Call of Duty game, the player can also choose their gender. With 20,000 possible combinations, this is the most flexible and comprehensive character customization in Call of Duty history. New dynamic maps are the evolution of multiplayer. They include interactive elements and player triggered events that make the environment evolve as each match goes on. The entire landscape can shift and force players to change tactics and strategies. Call of Duty: Ghosts introduces new tactical player movements. The new contextual lean system now allows players to lean around obstacles without adding button combinations or fully leaving cover. The new mantling system allows fluid movement over objects, while maintaining momentum. The knee slide allows for a natural transition from sprinting crouching to prone. Call of Duty: Ghosts delivers over 20 NEW Kill Streaks in Call of Duty: Ghosts - such as Juggernaut Maniac, the Helo Scout, the Vulture and the ODIN Strike. Players can even bring in guard dog Riley, from the single-player campaign, to protect and also to attack enemies. There are also over 30 NEW weapons, including an entirely new weapon class: Marksman Rifles.

#### **Squads**

Build your team and take up to 6 of them into battle in the all new Squads mode. This mode takes the best parts of the multiplayer experience and allows you to play either solo or cooperatively with the custom soldiers created and leveled up in multiplayer. The load-out choices you make for your Squad members will directly change the AI behavior of your squad-mates. Give your soldier a sniper rifle and he'll behave like a sniper, and an SMG guy will be more run and gun.



#### **Extinction**

A completely new one to four player, co-op game mode featuring a unique blend of fast-paced survival gameplay, base defense, player customization and class leveling. Teamwork and communication are essential when playing Extinction, as your team faces an inhuman menace that has overrun an isolated Colorado town and you must eliminate the threat via any means necessary.

source variable value	Attribute Modesofplay yes	Attribute Modesofplay no	Attribute Singleandmulti yes	Attribute Singleandmulti no	...	Consequence Creative yes	Consequence Creative no	Consequence Nostalgia yes	Consequence Nostalgia no
0	0	1	0	1	...	0	1	0	1
1	1	0	1	0	...	0	1	0	1
2	1	0	0	1	...	0	1	0	1
3	1	0	0	1	...	0	1	0	1
4	0	1	0	1	...	1	0	1	0
5	0	1	0	1	...	0	1	0	1
6	0	1	0	1	...	0	1	0	1
7	0	1	0	1	...	0	1	0	1
8	0	1	0	1	...	0	1	0	1
9	0	1	0	1	...	0	1	0	1
10	1	0	0	1	...	1	0	0	1
...	...	...	...	...	...	...	...	...	...
199	0	1	0	1	...	0	1	0	1

**Table 9:** Table showing a sample of the input data for MCA on 200 Video Game descriptions where there are two entries per observation (description) per attribute or consequence, indicating presence (1 in 'yes') or absence (1 in 'no') of that respective attribute or consequence in that observation.

## C. APPENDIX: ILLUSTRATIVE SAMPLE OF INPUT DATA FOR MCA ON 200 VIDEO GAME DESCRIPTIONS

source variable value	Attribute Modesofplay yes	Attribute Modesofplay no	Attribute Singleandmulti yes	Attribute Singleandmulti no	...	Consequence Creative yes	Consequence Creative no	Consequence Nostalgia yes	Consequence Nostalgia no
0	0	1	0	1	...	0	1	0	1
1	0	1	0	1	...	0	1	0	1
2	0	1	0	1	...	0	1	0	1
3	0	1	0	1	...	0	1	0	1
4	0	1	0	1	...	0	1	0	1
5	0	1	0	1	...	1	0	0	1
6	0	1	0	1	...	0	1	1	0
7	1	0	0	1	...	0	1	1	0
8	1	0	1	0	...	0	1	1	0
9	0	1	0	1	...	0	1	1	0
10	0	1	1	0	...	0	1	0	1
...	...	...	...	...	...	...	...	...	...
999	0	0	0	1	...	0	1	1	0

**Table 10:** Table showing a sample of the input data for MCA on 1000 Video Game reviews where there are two entries per observation (review) per attribute or consequence, indicating presence (1 in 'yes') or absence (1 in 'no') of that respective attribute or consequence in that observation.

#### D. APPENDIX: ILLUSTRATIVE SAMPLE OF INPUT DATA FOR MCA ON 1000 VIDEO GAME REVIEWS

*Für meinen Papili.*