



UvA-DARE (Digital Academic Repository)

Variability: The effects of variation in power relations within the firm, in its market performance, and in the evaluations of its products

Situmeang, F.B.I.

Publication date
2013

[Link to publication](#)

Citation for published version (APA):

Situmeang, F. B. I. (2013). *Variability: The effects of variation in power relations within the firm, in its market performance, and in the evaluations of its products*. [Thesis, fully internal, Universiteit van Amsterdam].

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

3. History matters: The impact of evaluations and sales of earlier versions of a product on consumer and expert evaluations of new editions of products²

Abstract

Product evaluations are assumed to be based on the observable characteristics of the underlying product. However, in the case of new editions of product the determinants may include signals that originate from the evaluations and the sales of editions that precede the focal product edition. Our analysis of 577 video games released between 2000 and 2009 indicates that the evaluations of earlier versions have a positive relationship with the evaluations of the sequel by the same type of evaluator. We also find that expert evaluations are influenced by the average evaluation of previous edition by consumers and the average sales of previous versions of the product. This suggests that experts tend to adapt to the taste of consumers. Furthermore, it is found that a lack of consensus, between evaluators of a particular type, weakens the impact of average past evaluations, whilst it magnifies the impact of the sales of earlier versions.

Keywords: Sequels, evaluations, experts, consumers, variability

3.1. Introduction

Consumer demand is highly uncertain in the creative industry, which is why producers tend to reuse ideas and concepts that have been known to be successful in the past (Tschang and Szczypula, 2006). As a result, the strategy of releasing new products as sequels has become increasingly popular in this industry (Rosen, 2011). New editions of successful brands such as the Harry Potter, Batman, Superman, and many others are brought to the market in the form of a product series. In the videogame industry, for example, the sequels of series such as Mario Brothers, Pokemon, The Sims, World of Warcraft, and Call of Duty, have generated billions of dollars of revenue (Philips, 2011).

The experiences that consumers have had with the editions that precede the sequel have been found to be the major contributor to the market success of these sequels (Völckner and Sattler, 2006). To explain this relationship, it has been argued that the ideas and

² This chapter is conditionally accepted in the European Journal of Management

impressions connected to the earlier editions are transferred to the next edition (Keller 2003; Erdem, 1998; Sunde and Brodie, 1993) and that the popularity of the past editions build up anticipation and excitement towards the new editions (Dhar, Sun, Weinberg, 2012; Ho, Dhar, and Weinberg, 2009; Karniouchina, 2011). In this way, the new editions can profit from the success of the earlier editions, but their market performance can also suffer from the comparison with the earlier success. The excitement and anticipation created by the previous performance generates high expectations that can easily lead to a lower level of satisfaction (Anderson, 1973; Oliver, 2009). It has also been found that the market performances of movie sequels are generally lower than that of the original editions (Basuroy and Chatterjee, 2008). Sequels by definition also retain the core attributes of their original parent products and are often considered less innovative (Sood and Dreze, 2006). However, this paper does not analyze the factors that predict sales, but the factors that predict the evaluations themselves.

Whilst previous studies have discussed the notion of image transfer from a product to a new edition, they appear to overlook how the positive appreciations of past editions in the product series are transferred to the evaluations of the new edition. Even though the notion of the product image being transferred from one edition to the next is widely accepted in the literature, signals such as product evaluations are usually treated as independent determinants of sales performance. At the same time, it is usually assumed that product evaluations (or reviews) are based on the observable characteristics of the underlying product. However, as we will argue more extensively below, in the case of new editions of products the determinants may be more complex and include peripheral signals that do not originate from the focal product itself, but reflect the performance of earlier editions.

The quality of a product is always the quality as perceived by observers, and sales performance as well as product evaluations provide indicators of how this quality is

perceived. When the core attributes remain the same between editions, it is reasonable to assume that the "underlying" quality of the product has not changed and that therefore evaluations of the new editions will not differ much from evaluations of the past editions, or just decrease a bit because of the same general reasons of unfulfilled expectations and lesser perceived innovativeness that have been proposed with regard to the sales performance of sequels. This paper looks more closely at the question whether evaluations of past editions indeed predict the evaluations of the new edition and proposes and tests arguments explaining why the evaluations of the new edition can be different from those of the past editions.

The first aim of the paper is to build on insights from a number of theoretical perspectives and, in particular, from signaling theory (Spence, 1973; Kirmani and Rao, 2000; Connelly et al., 2011) to construct and empirically test a model that explores the connections between how well past editions did, in terms of the evaluations of both consumers and experts (and of sales performance), on the expert and consumer evaluations of the new edition.

The second aim of this paper is to explore the effects of the lack of consensus in the evaluations of past editions as a determinant of the evaluations of future editions. Most earlier studies on the effects of product evaluations focus on the average evaluation, which hides the fact that consumer and expert evaluations are often polarized, leading to fluctuations around the mean (Sun, 2012). Consumers and experts can easily observe the differences and distribution of product evaluations (Chevalier and Mayzlin, 2006), and the dispersion or variability of evaluation have a significant effect to consumer behavior (Sun, 2012). Following this recent development in the literature, we include variability of product evaluations in our research. In particular, we explore the effects of the lack of consensus among evaluators of the same type evaluating the same edition - with regard to both communities of evaluators: the consumers and the experts.

Our study contrasts with earlier studies (e.g. Keller, 1993; Hennig-Thurau, Houston and Heitjans, 2009), in which consumer evaluations and expert critics evaluations were considered as items pertaining to the single construct "brand image". Other studies have also pointed out that consumers and professional experts are members of different communities that interact differently with competitive processes (Wijnberg and Gemser, 2000), and in which different criteria are used to determine product value (Bhansing, Leenders, and Wijnberg, 2012; Gemser, Leenders and Wijnberg, 2008; Plucker et al., 2009; Moon, Bergey, and Iacobucci, 2010). These studies suggest that, even if, in a particular industry, consumer and expert evaluations do have similar effects on sales, the determinants of the evaluations themselves - the main focus of this paper - can be very different.

The setting of the study is the creative industry, in particular, the video games industry. As we have pointed out before, evolving products through a series of new editions is a very popular strategy in this industry (Tschang and Szczypula, 2006), which makes this industry a suitable setting to our purpose. In addition, the creative industries offer the advantage of the availability of large publicly accessible online databases (e.g. metacritic.com, vgchartz.com, imdb.com, etc.) of evaluations and sales. Even though our analysis is based on the creative industry, the results of our analysis can be generalized to other industries where producers bring products on the market in a series and where product evaluations are influential factors for consumer buying behavior, such as the automotive industry, cellphones, or laptops among many others.

3.2. Theory and hypotheses

3.2.1. Product evaluations as signals and the conceptual model

Product evaluations, both by experts and other consumers transmit information about the product. When reliable information is scarce, especially if the quality of the product is difficult to evaluate before consumption, both types of evaluations can have a strong impact on consumer behavior (Basuroy, Chatterjee, and Ravid, 2003; Caves 2000). A recent study suggested that 62% of consumers read (online) evaluations and almost half of them (49%) read the evaluations of competing products/services before deciding which product or service to buy (Lightspeed Research, 2011). This supports an earlier study that found that 82% of consumers say that their purchase decisions have been directly influenced by product evaluations (Deloitte, 2007).

The prominence of evaluations in the online and offline world has led many scholars to explore how evaluations affect sales, with initial studies focusing on the impact of expert evaluations (e.g. Basuroy, Chatterjee, and Ravid 2003; Boatwright, Basuroy, and Kamakura, 2007; Eliashberg and Shugan, 1997; Reddy, Swaminathan, and Motley 1998). More recently there have been an increasing number of studies that pay attention to the effects of evaluations by consumers (e.g. Liu, 2006; Zhu and Zhang, 2010). Considered from the perspective of signaling theory (Spence, 1973; Kirmani and Rao, 2000; Connelly et al., 2011) both consumer evaluations and expert evaluations signal the quality of the product being evaluated, which in turn reduces consumers' uncertainties regarding the product.

Despite the interest in analyzing the effects of evaluations on sales, the determinants of the evaluations themselves have not been studied systematically. Rather it is assumed that the evaluations are simply reflections of how well the characteristics of the product fit quality standards and the preferences of the evaluator. It is assumed that the signals that are taken into account to arrive at the evaluation are those can be directly linked to the product that is

evaluated (Spence, 1973; Johnson and Levin, 1985; Kirmani and Rao, 2000). However, products in a series of editions may violate this assumption because each new edition strongly associated with previous editions in that series. For this reason, signals that pertain not to the product itself but to these previous editions - and their performance - can affect the evaluation of the focal product. The term "peripheral" signals is used to denote signals that do not directly relate to the characteristics of the evaluated product itself. Such peripheral signals include 1) the evaluations of earlier versions, and 2) the sales performance of earlier versions.

With regard to the effects of earlier product evaluations, the default expectation would be continuity; such that if the previous edition received on average a positive evaluation score from a community of evaluators, the focal edition will also be evaluated positively by that same community. However, average evaluations scores tell only part of the story. Another signal that is important is the variability of evaluations within each community of evaluators - experts or consumers - as this reflects the extent of the consensus among evaluators within their respective community. This can potentially influence the effects of past evaluations and past sales. The role of variability as an amplifier of peripheral signals on the evaluations of the focal edition will be discussed and tested more extensively in the following sections. The conceptual model that outlines the relationships of peripheral signals with expert evaluation and consumer evaluation is presented in Figure 3.1.

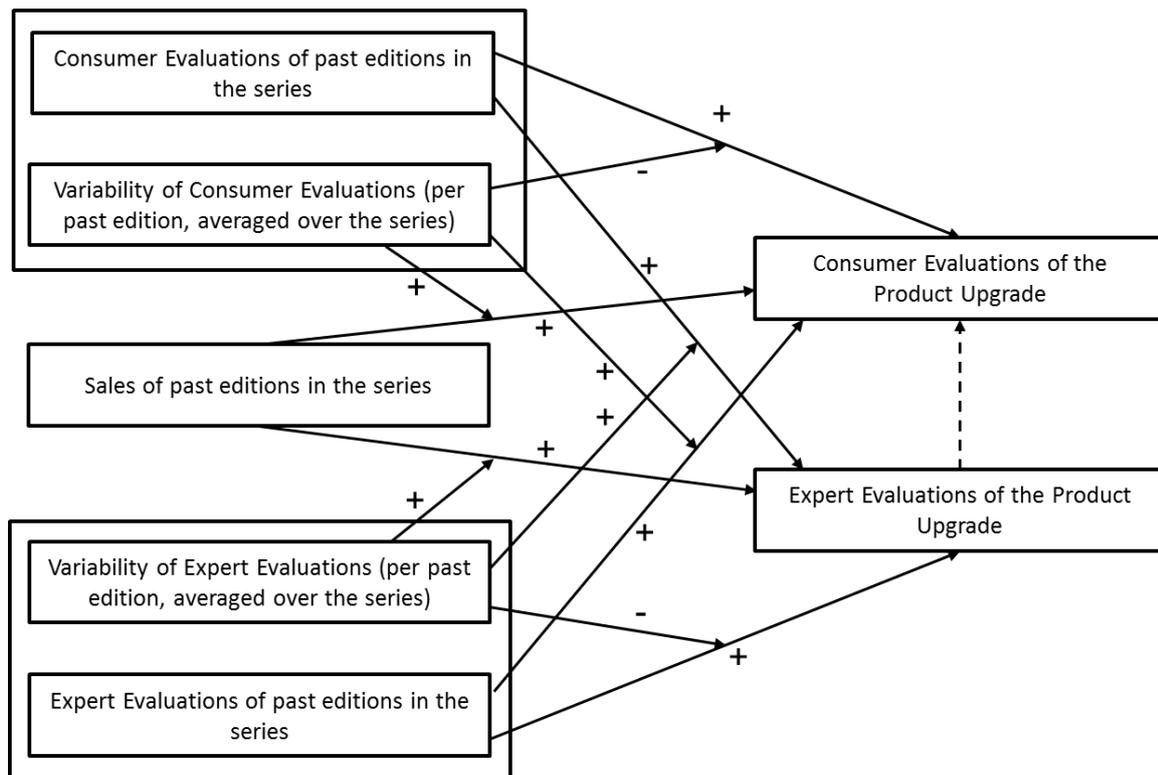


Figure 3.1 Conceptual Model of the Determinants of Expert and Consumer Evaluations of Product Upgrades

3.2.2. The relationship between past evaluations and the evaluation of new edition

Products with the same brand name are often closely linked to each other, and that information about one member of the series is also valid for another product from the same series (Erdem, 1998). The performance of new editions will be - at least partially - determined by the extent that there is a carry-over of the consumers' image of the past editions to the next editions (Keller 2003). In line with the previous research, past certifications of a baseball player has been found to influence his long-term reputation, by virtue of what is called the "reputation halo" (Griffin and Ward, 2010). If these arguments are seen from the perspective of signaling theory, past performance track records can signal the predicted future behavior (Perks and Halliday, 2003). In the same way, one may argue that past evaluations are also *signals* of consumers and experts appreciations of *the series*, thus

the evaluation affects not only the sales of the particular editions that the evaluations address, but also those of the potential new members of the series that are released *in the future*.

In addition to consumers' tendency to link past performance with current performance (Aaker and Keller, 1990; Erdem, 1998), consumers learn about and familiarize themselves with the "franchise" by consuming a series of sequels or editions of the product (Rothschild and Gaidis, 1981). The familiarity and the increased knowledge it implies, can be considered to be present in the community of consumers. Unless there is a major change in the product attributes, such knowledge is constantly reinforced by the characteristics of the "reminders" that conserve or elicit specific product memories (Hoyer, 1984). Therefore, the product evaluations by consumers of earlier editions in the same series signal are not just signals of how attractive to consumers these particular editions are, but also of the attractiveness of the franchise or series as a whole. Considering this aspect of consumer evaluations, as indicators of the quality, as perceived by consumers, of the series, one can expect there to be continuity in average response of consumers to the subsequent editions of the series. For this effect to operate, one does not have to assume that consumers know of the consumer evaluations of the previous editions. It suffices that these evaluations are at least partly influenced by characteristics that are attributed to the series as a whole. This suggests:

H1a: Consumer evaluations of past editions in the series have a positive relationship with consumer evaluations of the new edition of the product.

The general arguments presented above with respect to consumer evaluations seem also applicable to the relationship of expert evaluations of past editions with expert evaluations of the new edition. New editions or sequels are very similar to the original product in terms of core product attributes and physical form. Due to the similarity between

them, the attitudes towards the original product and the new edition can also be expected also similar (Aaker and Keller, 1990).

At the same time, there can be additional mechanisms in play with regard to the experts, which are not or much less relevant for the consumers. Expert reviewers might be more wary of publicly changing their opinions about particular franchises, since their professional reputation will be linked to their evaluations and changing their mind might harm their credibility. Consequently, expert opinions of the new edition would be very similar to the evaluations given to the earlier products in the series – unless the new edition deviates significantly from the core attributes and characteristics of the series.

The above argument suggests that the expert evaluations will remain even more constant than those of the consumers. However, there is also an argument suggesting an effect in the opposite direction. Professional critics are more likely to notice in new editions of the same product a lack of creativity and innovation (e.g. Hudson, 2011; Kondoloy, 2012), which suggests that maintaining a similar product attributes would lead to a downward trend of expert evaluation from one edition to the next. This is especially likely in high-art industries, in which innovativeness as such completely dominates the quality judgment (Wijnberg and Gemser, 2000), while in other industries precisely maintaining the positively-valued traditional attributes of the brand or franchise is more likely to be approved of. Thus, we would not expect the additional (compared to the case of the consumers) negative effect to be very strong or outweigh the possible additional positive effect. It is therefore hypothesized that:

H1b: Expert evaluations of past editions in the series have a positive relationship with expert evaluations of the new edition of the product.

3.2.3. The relationship between past sales performance and the evaluation of new edition

Sales performance of a previous edition denotes the number of units of a product sold to consumers, which is also an indicator of the attractiveness of the previous version among the consumers (Hennig-Thurau, Houston and Heitjans, 2009). As one main reason of a signal is to reduce the uncertainty about a product by informing consumers what can be expected from it, past sales as a peripheral signal can also serve as a reflection, to some extent, on how well the series has been appreciated over time by consumers in general across past editions in the series.

In addition to signaling the popularity of the product among consumers, past sales also indicate how many consumers know about the product series. The higher the number of consumers who know about the series, the higher the likelihood of Word-of-Mouth communication about the product, which in itself can have an effect on subsequent sales (Duan, Gu and Whinston, 2008). Consumers talk more about financially successful movies than unsuccessful movies, which would affect subsequent revenues (Moon, Bergey and Iacobucci, 2010). So, the larger the number of product evaluations from individual consumers, the more likely it will be that consumers will communicate with other consumers about the product, reinforcing possible positive sentiments towards the product, as well as creating social pressure to try the new edition. It is therefore hypothesized that:

H2a: The sales performance of past editions in the series has a positive relationship with consumer evaluations of the new edition of the product.

It may be argued that past sales are less likely to influence critics. Consumer evaluations and sales performance are both measurements of how well consumers appreciate the series. Expert critics, however, are supposedly different from consumers. Precisely

because they are professionals, it is commonly assumed that expert critics are *independent* evaluators, who often, at least in the case of movies, make up their opinions before the public launch of the product (Debenedetti, 2006). Some might be susceptible to bribes of favors from the producers (Cameron, 1995; Eliashberg and Shugan, 1997; Mol and Wijnberg, 2007), but sales performance in itself is not likely to sway their judgment. They are assumed to be unbiased, willing to educate the public, and to evaluate based on how they perceive the product, rather than being influenced by the success of earlier editions of the series among consumers.

However, the real state of affairs may not match this idealized picture, particularly in market driven industries such as the video games industry (Adams and Rollings, 2003; Phillips, 2011). The dominant selectors in a market driven industry are the consumers (Wijnberg and Gemser, 2000). Professional experts are selected by the media in which their evaluations appear (Liu, 2006; Chen, Liu and Zhang, 2012). They are often employed by, or at least affiliated to, media outlets, i.e. newspapers, magazines, tabloids, radio/television shows, websites, etc., and obtain economic compensations for the reviews that they write. Most of these outlets are profit-seeking entities and are prone to market pressure in terms of maintaining readership to ensure their economic viability. The media outlets tend to the needs of their readers and elevate them as 'the newsroom's giants with the status of equal partner in decision making' of the media's content (Beam, 2003, 368). In this way, the perceived preferences of the readers are likely to influence the judgments of the expert reviewers who work for the media. In so far as readers might also be consumers of the product, the sales success of earlier editions might provide a strong signal to the media outlets, and then, directly or indirectly, to the expert critics themselves, that this is a product that readers like. The first effect of observed past popularity on reviews would be to increase the chance of being reviewed. This is of particular importance in industries where only a fraction of

products is professionally reviewed, such as book publishing. Furthermore, one can expect a strong effect on the valence of the reviews, and this is of relatively greater importance in an industry where most or all products can expect to receive expert reviews, such as mainstream movies or video games. When there is a clear signal that a series is particularly popular among the consumers, writing a negative review about a new edition in this series means that the expert writes against the 'will' of the market. By doing so, the expert would risk estranging him/herself, and the medium in which the review appears, from the readers, particularly in industries where market selection is very dominant. In addition, if an expert's opinion is not in line to the consumers' preferences, the readers can also start to question the credibility of that expert. Therefore, if the product is part of a series, earlier sales performance could well have an effect on the expert reviews. This suggests:

H2b: Sales performance of past editions in the series has a positive relationship with expert evaluations of the new edition of the product.

3.2.4. Consensus in the community of evaluators as a signal-amplifier

In most social groups evaluations of the same object will vary among group members due to individual differences, in taste or perspective (Boor, 1990; West and Broniarczyk, 1998). A high variability of the evaluations of the same object within the same group can be understood as a lack of consensus, among group members, with regard to the object. This variability of opinions has received little attention in the literature on the effects of evaluations (exceptions are: Sun, 2012; and Das and Chen, 2007).

Initially, the discussion regarding the effect of variability of opinions focused on the finance sector (Das and Chen, 2007). The results of this research suggest that there is a strong relation between the variability of opinions among stockbrokers (as observed in their web

discussions) and downward movements in the actual stock price (Das and Chen, 2007). More recently, it is also found that the interaction of the average consumer evaluations of books and the standard deviation of these evaluations has a significant effect on the demand, i.e. products with low ratings would benefit from high variability of ratings, while products with higher rating would suffer from the variability of the ratings (Sun, 2012). In contrast to the previous two studies, the main aim of our research is to study the determinants of evaluations, not of demand. Moreover, we will look at variability as a moderator of the effects of the other signals, such as average past evaluations.

We start again from the perspective of the consumer. With regard to each past edition of the product there can be a higher or lower consensus. If, on average, consensus is low over the series of past editions, this can signal that, on an aggregate level, it was very difficult for the evaluators as a group to assess the qualities of these past editions. Although expert evaluators may be more likely to pay explicit attention, consumers do as well observe and notice the differences and the distributions of product evaluations across time (Chevalier and Mayzlin, 2006; Sun, 2012). Given that, as we argued above, average evaluations of past editions will have an effect as a signal of quality pertaining to the franchise or the series as a whole, it is likely that high variability in these evaluations of past editions will diminish the effect of this signal, precisely because the high variability, and especially variability that remains high over a series of editions, will denote that the evaluators in a particular community cannot easily agree about the quality.

Additionally, a lack of consensus decreases the chance that consumers will expect to see their opinions reinforced by other consumers, which may affect the likelihood that they will express their opinions or, at least, express them strongly (c.f. Neuwirth, Frederick, and Mayo, 2007; Oshagan, 1996). These arguments suggest:

H3a: The relationship of past consumer evaluations with consumer evaluations of the new edition of the product is moderated by the degree of consensus among consumer evaluators of past editions. The relationship is less positive if there is a lower degree of consensus.

With regard to the experts, we would expect a similar general effect, that the evaluations of past editions become less convincing as a signal of the qualities of the whole series when consensus among expert evaluators of these past editions has been low. Moreover, given that expert critics are professionals, they often find it necessary - also as a sign of professionalism - to take their peers' opinions into account and even comment on them in their reviews (Das and Chen, 2007). Although in high art industries there might be a premium on being an innovator among reviewers (Wijnberg and Gemser, 2000), in most other contexts a striving for consensus among critics has been observed (Van Rees, 1987; Hsu, 2006; Shoemaker and Vos, 2009), eventually leading to consecration of some products (Allen and Lincoln, 2004; Allen and Parson, 2006; Graffin and Ward, 2010). When experts generally agree, there will be very little benefit for an individual expert in publicly disagreeing. However when there is no clear consensus among the members of the expert community, individual experts will be less constrained and more likely to express opinions that diverge from the average expert evaluations of past editions. This suggests:

H3b: The relationship of past expert evaluations with expert evaluations of the new edition of the product is moderated by the degree of consensus among expert evaluators of past editions. The relationship is less positive if there is a lower degree of consensus.

As has previously noted, the main function of evaluations as signals is to reduce uncertainties by providing information regarding the product (Kirmani and Rao, 2000). If particular signals leave the consumers with an ambiguous message and increasing uncertainty, consumers will rely more strongly on other types of information (Choi, 1993). The lack of consensus among reviewers would lead to consumers seeking alternative sources of information to form estimates of the quality of the product (West and Broniarczyk, 1998). Thus, when a particular signal creates more uncertainties, other signals will be more likely to have an effect on consumer behavior. As discussed, a lack of consensus among the evaluators of a particular type makes the average evaluations of past editions less clear or trustworthy signals to the evaluators, of the same type, of new editions.

When consumers evaluations of past editions are highly variable, consumers will give more weight to other signals, and the same will hold for consumer evaluators. As discussed, the sales of previous editions are also signals of the quality of the whole series to the consumer evaluators. We would therefore expect that the weaker the signal of the evaluations of past editions is, because of the low consensus, the more the consumer evaluators will rely on the other signal conveying similar content, namely sales of past editions.

H4a: The relationship of past sales with consumer evaluations of the new edition of the product is moderated by the degree of consensus among consumer evaluators of past editions. The relationship is more positive if there is a lower degree of consensus.

Again, we can start with a similar argument for the experts. Lower consensus among expert evaluators of past editions can be expected to increase the strength of the signal originating from the sales performance of past editions. It could even be expected that the increase in strength, due to the variability of evaluations of past editions, may be greater with

regard to the experts than to the consumers. If expert evaluators are cautious about producing opinions that are less relevant to their audience, they will be even more likely to disregard other signals if they are not completely convincing, and instead pay full attention to sales of past editions. It is therefore hypothesized that:

H4b: The relationship of past sales with expert evaluations of the new edition of the product is moderated by the degree of consensus among expert evaluators of the past editions. The relationship is more positive if there is a lower degree of consensus.

3.3. Research method

3.3.1 The sample

To test the relationships of the peripheral signals to the evaluations of new editions, we benefit from online data repository that provides extensive performance records of creative products: vgchartz.com for the sales and metacritic.com for the evaluations. These online databases have also been used in earlier studies (e.g. Chen, Liu, and Zhang, 2012; Hennig-Thurau, Houston, and Heitjans, 2009; Zhu and Zhang, 2010).

The performance records belonging to 212 series released between 2000 and 2009, $n = 577$, have been collected from these websites. The dataset covers video games sequels released on all three major consoles, i.e. Nintendo Wii, Microsoft Xbox, and Sony PlayStations within the time period. The average number of editions per series in the sample is 3 (excluding the original), and the largest is 10: sporting games had the highest average number (6) of editions per series.

3.3.2 The operationalization of the variables

Expert Evaluations and *Consumer Evaluations* data are obtained from the Metacritic's (www.metacritic.com) database of video games, which has been used in many previous studies (c.f. Hennig-Thurau, Houston, and Heitjans, 2009; Plucker et al., 2009). For each type of evaluation, the average of past evaluations in the series is measured by averaging relevant average evaluation scores of past editions in the series. For example, the average consumer evaluation of past editions in the series is measured by averaging the average consumer evaluation scores per edition in the series before the focal product edition.

The *Lack of Consensus* denotes the variability of evaluations given by the evaluators of one community to each edition in the series. There are of course various way to measure variability, such as range, interquartile range, or mean difference. However, variance is the method of choice because it is the most commonly used measure of variability in the marketing literature (e.g. Dacin and Smith, 1994; Dwyer and Welsh, 1985). For each of the editions before product edition i in series j , the variance of the evaluation scores is measured and then the arithmetic mean of the calculated variances of each edition prior to the focal product is calculated, giving the measure of the past variability of the product evaluation.

$$Variability_{i,j} = \frac{1}{i-1} \cdot \sum_{n=1}^{i-1} [Var(Evaluation)_{n,j}]$$

Sales Performance data are obtained from the database of vgchartz.com. This website shows the number of legal copies sold worldwide. This approach makes sense because console gaming requires users to have the discs to play; so - assuming second hand re-sales and illegal copying are proportional to legal sales – the number of copies sold is a good indicator of the size of the user community. Just like the average of past evaluations, the average past sales are measured by calculating the average of worldwide sales of all editions in the series before the focal sequel. The average sales across past editions is used because

consumers may skip some releases in the series, therefore, the construct can capture the overall number of potential consumers of the new edition based on the overall sales performance of the series. A further benefit of this approach is that the sales figures do not have to be corrected for inflation or changes in the exchange rates. It is acknowledged that the sales measure excludes digital video game download, which is very common for the PC platform. However, this type of product distribution is nonexistent in the console platforms studied in this paper.

This study uses the change of video games genre, change of video game age rating, and change of product developer of the new edition, to control for possible attribute changes that may influence consumer or expert evaluations of the new editions. This study also controls for a range of variables, namely: platform manufacturer and how many editions there have been before the focal sequel.

3.4. Empirical results

3.4.1. Description of the Data

Table 3.1 presents the correlation matrix and the descriptive statistics of the variables used in this research. On average sequels in the sample received positive evaluations from both experts and critics, both evaluation scores average above 7 out of 10. On the contrary, it seems that expert evaluations ($\mu = 4.96$) display a larger variability than consumer evaluations ($\mu = 3.75$), even though there are fewer experts ($\mu = 24.65$) than consumers ($\mu = 44.71$) who evaluated the products in the dataset. This is an early indicator that there may be different dynamics at work in the two communities.

The results of the correlation analysis can be used as preliminary evidence to support the basis hypotheses. Already in this test, it is observed that evaluation scores of the new

editions are highly correlated with the respective community past evaluations on the series ($r = .269$, $p < .001$ for the consumers, $r = .319$, $p < .001$ for the experts). In addition, there is a significant and positive relationship between past sales and consumer evaluations ($r = .107$, $p < .05$) and unexpectedly an even stronger relationship with the expert evaluations ($r = .226$, $p < .001$). Most importantly, there is no correlation above .5 between the independent variables; therefore, the independence of the constructs is verified.

Table 3.1 Descriptive Statistics and Correlation Matrix

| | Mean | SD | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|--------|---------|----------|----------|----------|-------|---------|----------|----------|----------|-------|----------|-------|------|
| 1 Consumer Evaluations of the Sequel | 7.814 | 1.491 | | | | | | | | | | | | |
| 2 Expert Evaluations of the Sequel | 7.437 | 1.257 | .353*** | | | | | | | | | | | |
| 3 Average Consumer Evaluations of Past Editions (PE) | 8.021 | 1.086 | .269*** | .085 | | | | | | | | | | |
| 4 Average Variance of Consumer Evaluations of PE | 3.746 | 4.145 | .010 | -.038 | .057 | | | | | | | | | |
| 5 Average Quantity of Consumer Evaluators of PE | 44.711 | 135.624 | .127*** | -.048 | .121*** | -.020 | | | | | | | | |
| 6 Average Expert Evaluations of PE | 7.198 | 2.150 | .103* | .319*** | .253*** | -.015 | .163*** | | | | | | | |
| 7 Average Variance of Expert Evaluations of PE | 4.963 | 6.522 | -.072 | -.140*** | -.152*** | -.008 | -.104** | -.224*** | | | | | | |
| 8 Average Quantity of Expert Evaluators of PE | 24.654 | 18.511 | .143*** | .171*** | .280*** | -.054 | .463*** | .193*** | -.341*** | | | | | |
| 9 Average Sales of PE | 1.638 | 1.836 | .107* | .226*** | .074 | -.041 | .536*** | .277*** | -.169*** | .404*** | | | | |
| 10 Number of Editions in the Series | 3.459 | 3.789 | -.144*** | -.119*** | -.080 | -.055 | -.088 | -.221*** | -.050 | -.215*** | .028 | | | |
| 11 Changes of Age Restriction (dummy) | 0.098 | 0.297 | .074 | .110* | .089 | -.019 | -.015 | .020 | .040 | .037 | -.001 | -.066 | | |
| 12 Changes of Genre (dummy) | 0.178 | 0.383 | -.061 | -.117*** | -.105** | -.020 | -.064 | -.116** | .089 | .012 | -.001 | -.138*** | .019 | |
| 13 Changes of Developer (dummy) | 0.161 | 0.368 | -.070 | -.116** | -.050 | -.018 | -.003 | -.082 | .062 | -.064 | -.054 | .120*** | -.025 | .028 |

*** significant at < .001

** significant at <.01

* significant at <.05

3.4.2. Pre-test: endogeneity

Assuming that video games developers are profit seeking entities, they would discontinue unsuccessful product series and this may create endogeneity problems that need to be addressed. Ho *et al.* (2009) suggest a two stage equation system to tackle a similar endogeneity problem where they predict the effect of showing an advertisement (trailer) of a movie during the Super Bowl halftime break. They first predict the probability of airing a trailer and the estimated probability is used as a predictor in the main model. In a similar fashion, there could be several factors that influence management decisions to bring a new edition in a series to the market, including performance of past editions. To tackle this issue, we develop a probabilistic function that estimate the probability of sequel i being released to market based on instrumental variables. We use the market performance of only *one* edition immediately before the sequel i . Even if there is a series effect, the sales between editions can vary highly. Besides sales, market condition can vary across editions. For this we include Competition (COMP) as the next instrumental variable. This variable is measured by the number of other games released in the same year that have the same genre attribute. In addition, these control variables are also included in the probabilistic model: Multiplayer or Single player (MULTI), ESRB age restriction (Teen and Mature; Everyone is the default and excluded from the model), and Genre classifications (First Person Shooter (FPS), Racing, Role Playing Games (RPG), and Sports). The function is mathematically written as follows:

$$\begin{aligned} & \text{Probability}(\text{releasing a new sequel } i) \\ &= \Phi(\alpha_0 + \alpha_1 \cdot \text{SALES}_{i-1} + \alpha_2 \cdot \text{COMP}_{i-1} + \alpha_3 \cdot \text{MULTI}_{i-1} + \alpha_4 \cdot \text{TEEN}_{i-1} \\ &+ \alpha_5 \cdot \text{MATURE}_{i-1} + \alpha_6 \cdot \text{FPS}_{i-1} + \alpha_7 \cdot \text{RACING}_{i-1} + \alpha_8 \cdot \text{RPG}_{i-1} \\ &+ \alpha_9 \cdot \text{SPORTS}_{i-1} + \varepsilon_i) \end{aligned}$$

This model (and simpler models not using the Sales (i-1) variable, which resulted in similar results) is estimated using a probit function, and then the resulting probabilistic scores are included as an extra variable ($\hat{p}(\text{Sequel})_i$) in our main model as a predictor of the consumer

evaluation and expert evaluation of the sequel (Ho, Dhar, and Weinberg, 2009).

3.4.3. Test of hypotheses

The hypotheses are tested using a path model analysis with Maximum Likelihood Estimation using AMOS (Arbuckle, 2009). Mathematically, the model can be written as the followings^{3,4}:

$$\begin{aligned} ConsEval_i = & \alpha + \beta_1 AvgConsEvalPE_i + \beta_2 AvgExpertEvalPE_i + \beta_3 VarConsEvalPE_i \\ & + \beta_4 AvgSalesPE_i + \beta_5 AvgConsEvalPE * VarConsEvalPE_i \\ & + \beta_6 AvgSalesPE * VarConsEvalPE_i + \beta_7 AvgExpertEvalPE \\ & * VarConsEvalPE_i + \beta_8 ChangeAge_i + \beta_9 ChangeGenre_i \\ & + \beta_{10} ChangeDev_i + \beta_{11} NumberEdition_i + \beta_{12} Nintendo_i + \beta_{13} Sony_i \\ & + \beta_{14} ExpertEval_i + \beta_{15} \hat{p}(Sequel)_i + \varepsilon_i \end{aligned}$$

$$\begin{aligned} ExpertEval_i = & \gamma + \vartheta_1 AvgConsEvalPE_i + \vartheta_2 AvgExpertEvalPE_i \\ & + \vartheta_3 VarExpertEvalPE_i + \vartheta_4 AvgSalesPE_i + \vartheta_5 AvgExpertEvalPE \\ & * VarExpertEvalPE_i + \vartheta_6 AvgSalesPE * VarExpertEvalPE_i \\ & + \vartheta_7 AvgConsEvalPE * VarExpertEvalPE_i + \vartheta_8 ChangeAge_i \\ & + \vartheta_9 ChangeGenre_i + \vartheta_{10} ChangeDev_i + \vartheta_{11} NumberEdition_i \\ & + \vartheta_{12} Nintendo_i + \vartheta_{13} Sony_i + \vartheta_{14} \hat{p}(Sequel)_i + \omega_i \end{aligned}$$

The fit of the model is sufficiently high ($\chi^2=8.512$, $p = .385$, $RMSEA = .013$; $CFI = .981$; $GFI = .978$). The squared multiple correlations (R^2) of the dependent variables are .201 for consumer evaluation and .297 for expert evaluation. The results are reported in Table 3.2.

³ c.f. Byrne, 2000: 11

⁴ We also estimated a mixed model in which we allowed α and γ to vary across series. This takes into account the series-specific effects. The results, however, did not significantly differ from our AMOS results. We did not find significant variations of the intercept across series for the consumer model, i.e. the standard deviation of the intercept was 0. There was a slight variation in the intercept regarding expert evaluations across different series (the sigma of the intercept was .308(.082)), however, all the coefficients, significances and signs, were very similar. For example, the interaction between expert variability and sales was Coeff. = .140 (.013) in the mixed model, while it was Coeff. = .122(.010) in our original AMOS model.

Table 3.2 Standardized Path Coefficients

| | β | Sig. | Hypothesized effect |
|--|---------|-------|---------------------|
| Simple Effects | | | |
| Avg. Cons. Eval. Past Editions (PE) \rightarrow Cons. Eval. Sequel (S) | .219 | <.001 | + |
| Avg. Expert Eval. PE \rightarrow Expert Eval. S | .198 | <.001 | + |
| Avg. Expert Eval. PE \rightarrow Cons. Eval. S | .219 | <.001 | N/A |
| Avg. Cons. Eval. PE \rightarrow Expert Eval. S | .186 | <.001 | N/A |
| Avg. Sales PE \rightarrow Cons. Eval. S | .166 | .003 | + |
| Avg. Sales PE \rightarrow Expert Eval. S | .100 | .038 | + |
| Avg. Var. of Cons. Eval. PE \rightarrow Cons. Eval. S | -.095 | .066 | N/A |
| Avg. Var. of Expert Eval. PE \rightarrow Expert Eval. S | -.386 | <.001 | N/A |
| Interaction Effects | | | |
| Avg. Cons. Eval. PE * Avg. Var. of Cons. Eval. PE \rightarrow Cons. Eval. S | -.127 | .008 | - |
| Avg. Expert Eval. PE * Avg. Var. of Expert Eval. PE \rightarrow Expert Eval. S | -.144 | .005 | - |
| Avg. Expert Eval. PE * Avg. Var. of Cons. Eval. PE \rightarrow Cons. Eval. S | .021 | .705 | N/A |
| Avg. Sales PE * Avg. Var. of Cons. Eval. PE \rightarrow Cons. Eval. S | .102 | .043 | + |
| Avg. Cons. Eval. PE * Avg. Var. of Expert Eval. PE \rightarrow Expert Eval. S | .038 | .309 | N/A |
| Avg. Sales PE * Avg. Var. of Expert Eval. PE \rightarrow Expert Eval. S | .122 | .010 | + |

Sig. levels are two tailed

The results of the analysis of our model suggest that consumer evaluations of the new editions are positively related to the average consumer evaluations across past editions ($\beta = .219$, $p < .001$), average expert evaluation across past editions ($\beta = .219$, $p < .001$), and average sales across past editions ($\beta = .166$, $p < .01$). The positive relationship of the average past consumer evaluation is significantly lowered ($\beta = -.127$, $p < .01$) by the lack of consensus between consumers. This relationship is visualized in Figure 3.2. We also find evidence to suggest that the positive relationship of average past sales can be increased by the lack of consumer consensus ($\beta = .102$, $p < .05$), which is reflected by Figure 3.3. These results offer support for H1a, H2a, H3a, and H4a.

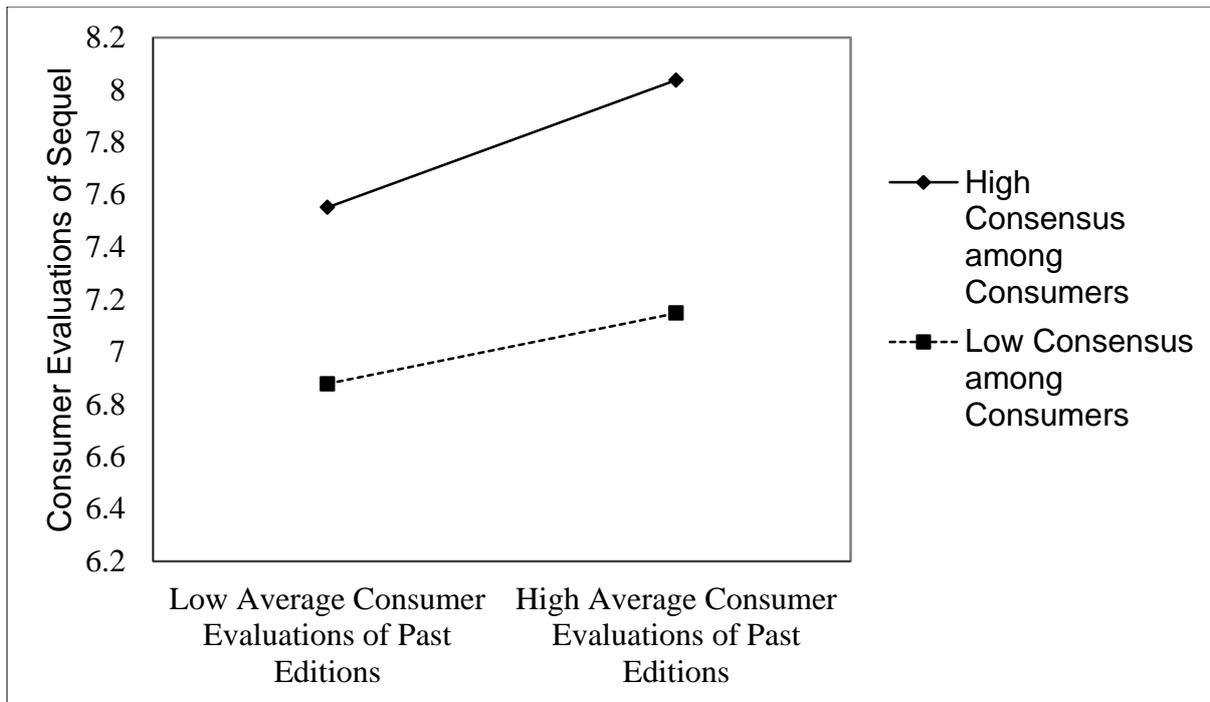


Figure 3.2 The effect of consumer evaluations of past editions on consumer evaluations of the product upgrade with low or high variability of consumer evaluations

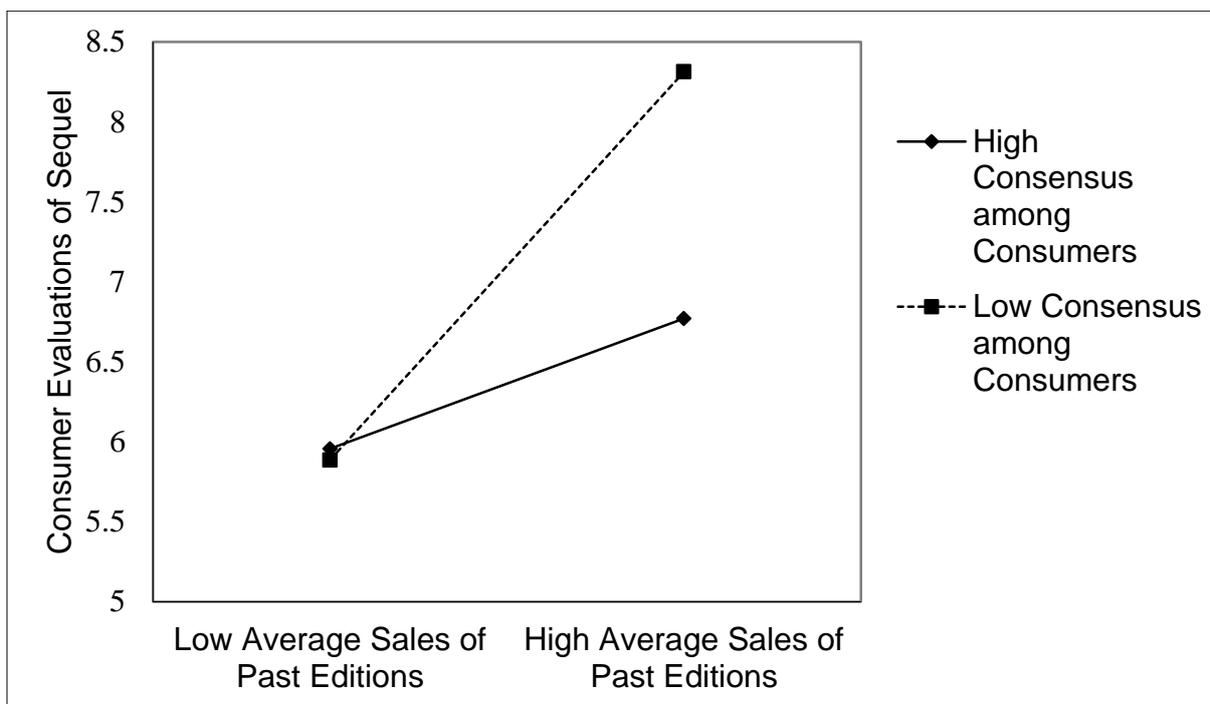


Figure 3.3 The effect of sales of past editions on consumer evaluations of the product upgrade with low or high variability of consumer evaluations

The results also provide strong evidence that expert evaluations of the new editions

are highly affected by average expert evaluations across past editions ($\beta = .198, p < .001$) and average sales across past editions ($\beta = .100, p < .001$). Besides that, we also find significant cross-over relationships, i.e. relationship of average consumer evaluations across past editions with expert evaluations of the new edition ($\beta = .186, p < .001$). The positive effect of past expert evaluations across editions is significantly lowered ($\beta = -.144, p < .01$) by the lack of consensus between the experts (visualized in Figure 3.4). Furthermore, we find strong evidence that effect size of past sales is more positive when there was lack of consensus among experts ($\beta = .122, p < .05$). This amplifying effect is visualized in Figure 4. With the previously mentioned results, H1b, H2b, H3b, and H4b are also supported.

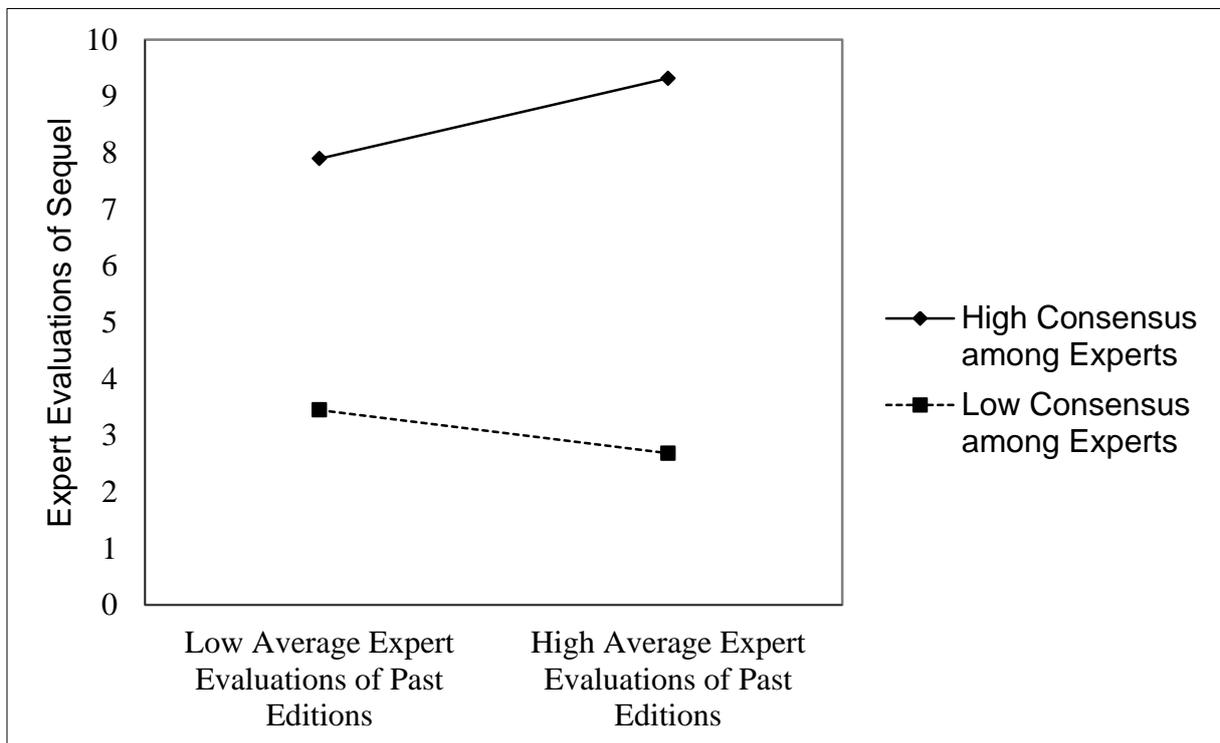


Figure 3.4 The effect of expert evaluations of past editions on expert evaluations of the product upgrade with low or high variability of expert evaluations

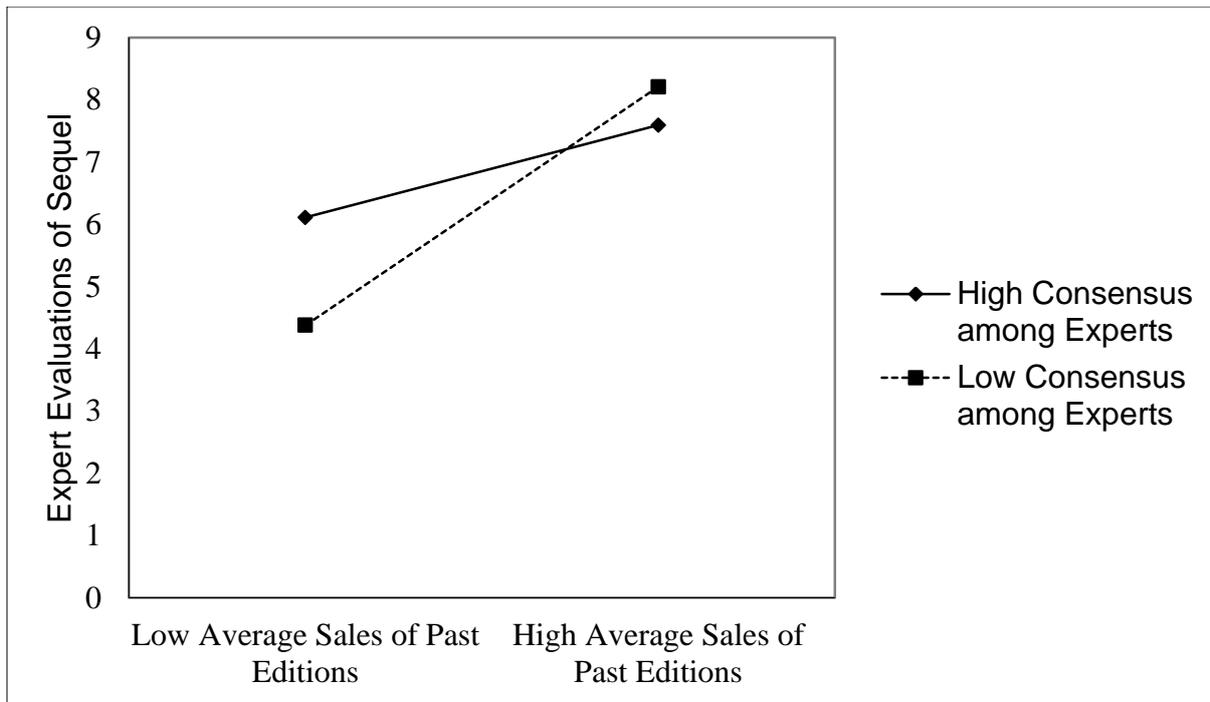


Figure 3.5 The effect of sales of past editions on expert evaluations of the product upgrade with low and high variability of expert evaluations

Consistent with the results of earlier studies on sequels, we found a negative effect of the number of editions in the series on product evaluation (for consumers $\beta = -.111$, $p < .01$, and for experts $\beta = -.109$, $p < .01$). The default platform during the model estimation is Microsoft Xbox games. There are no significant differences between the evaluations of sequels of Xbox and of Nintendo games by experts ($\beta = .003$, $p > .1$) or consumers ($\beta = .011$, $p > .1$). Experts evaluations are not significantly different for PlayStation games ($\beta = .052$, $p > .1$), though consumer evaluations are somewhat higher ($\beta = .139$, $p < .001$). Regarding the effect of changes of core product attributes, surprisingly, only 'change of genre' results in a significant effect on new product edition evaluations, and this effect is only limited to expert evaluations ($\beta = -.113$, $p < .01$). The significance level of the other control variables is higher than .10, suggesting that their impact on the expert and consumer evaluations is negligible.

3.4.4. Robustness tests

It may be difficult to comply with the multivariate normality assumption if one uses a dataset from the creative industry. For example, Hu, Pavlou, and Zhang (2006) showed that online book reviews are negatively skewed, which violates the standard normal distribution assumption of a dependent variable, and argued that this is because consumers' main motivation to write an online review is because of their satisfaction with the product. Similar to the findings in earlier studies, consumers evaluations in the sample used in this study are also skewed to the same direction (skewedness = -6.970). To deal with this deviation from the normality assumption, bootstrap resampling is performed to create a much larger sample and correct the bias of standard errors of the estimates (Arbuckle, 2009; Nevitt and Hancock, 2001). The bootstrap estimation is based on ML method with 2,000 times resampling. Similar significance levels of the estimates were reported for every coefficient, confirming the results that were presented in the main analysis.

We performed a 3SLS as an alternative simultaneous estimation procedure to estimate the determinants of evaluations. We choose 3SLS as an alternative because similar to AMOS Path Modeling that we adopted in our main analysis, 3SLS is a form of statistical model that simultaneously estimate a set of linear equations (Greene, 2003). The results are presented in Table 3.3. and they are very similar to the initial findings. The 3SLS Multiple R-squares are .230 and .254 respectively for the consumers and experts evaluations equations. Another way to test the robustness of the findings is by using a different way of measuring the variables tested in the model. As an alternative, this study follows the methodology implemented by Hennig-Thurau, Houston and Heitjans (2009) which used the performance of the edition immediately before the focal edition as the determinant of performance. Thus, instead of past sales, past evaluation scores, and the lack of consensus in past evaluations over the whole series of past editions, we can look only at these variables with regard to the

last edition before the focal one. The same model is estimated with the new variables, and the results (also presented in Table 3) show that most of the relationships are similar.

Table 3.3 Robustness test applying alternative methodologies of estimation

| | 3SLS (all past editions before the product upgrade) | | AMOS (but only 1 edition before the product upgrade) | |
|--|---|-------|--|-------|
| | B | Sig. | β | Sig. |
| Simple Effects | | | | |
| Avg. Cons. Eval. Past Editions (PE) → Cons. Eval. Sequel (S) | .357 | <.001 | .181 | <.001 |
| Avg. Expert Eval. PE → Expert Eval. S | .381 | <.001 | .261 | <.001 |
| Avg. Expert Eval. PE → Cons. Eval. S | .071 | .218 | .008 | .883 |
| Avg. Cons. Eval. PE → Expert Eval. S | .138 | .008 | .071 | .164 |
| Avg. Sales PE → Cons. Eval. S | .106 | .052 | .136 | .004 |
| Avg. Sales PE → Expert Eval. S | .119 | .001 | .182 | .001 |
| Avg. Var. of Cons. Eval. PE → Cons. Eval. S | -.005 | .489 | -.022 | .710 |
| Avg. Var. of Expert Eval. PE → Expert Eval. S | -.096 | .002 | -.432 | <.001 |
| Interaction Effects | | | | |
| Avg. Cons. Eval. PE * Avg. Var. of Cons. Eval. PE → Cons. Eval. S | -.164 | .006 | -.109 | .036 |
| Avg. Expert Eval. PE * Avg. Var. of Expert Eval. PE → Expert Eval. S | -.334 | <.001 | -.165 | .017 |
| Avg. Expert Eval. PE * Avg. Var. of Cons. Eval. PE → Cons. Eval. S | .010 | .459 | .039 | .471 |
| Avg. Sales PE * Avg. Var. of Cons. Eval. PE → Cons. Eval. S | .005 | .891 | .014 | .794 |
| Avg. Cons. Eval. PE * Avg. Var. of Expert Eval. PE → Expert Eval. S | .027 | .702 | .018 | .763 |
| Avg. Sales PE * Avg. Var. of Expert Eval. PE → Expert Eval. S | .098 | .017 | .053 | .244 |

Sig. levels are two tailed

Next, instead of using the evaluation scores, the number of consumers or experts who evaluate the products is used as an alternative measure of interest by consumers and experts (c.f. Gemser, van Oostrum and Leenders, 2007). Nevertheless, this study can find support for most of the effects presented in the main results. The number of reviewers of the new editions are highly affected by the average number of reviewers of the past editions (for consumers $\beta = .275$, $p < .001$, for experts $\beta = .577$, $p < .001$), and these effects are mitigated by the lack of consensus in the respective communities (for consumers $\beta = -.258$, $p < .001$, for experts $\beta = -$

.485, $p < .001$). However, there is no significant effect of past sales on the number of consumer evaluations of new editions ($\beta = .064$, $p = .214$) or on the number of expert evaluations ($\beta = .062$, $p = .199$). It seems likely that, because of strong relationship between sales and number of evaluators, the effects of the past sales are absorbed in the effects of the number of past evaluators. Interestingly, the explanatory powers of the model (R^2) also increase ($R^2 = .306$ for consumers and $R^2 = .386$ for experts).

As an alternative to the average variability among evaluation scores within the editions, the variability of the average evaluations across editions is used to measure the lack of consensus, please refer to the following formula:

$$\text{Variability_Alternative}_{i,j} = \text{Var}[\text{Mean}(\text{Evaluation})_{0\dots i-1,j}]$$

Arguably, the two are conceptually different factors; the first measures the degree of the lack of consensus about a single object of evaluation in a community of evaluators, while the latter measures the fluctuation of perceived product quality across editions in a series. However, both variables signal uncertainty about product quality (c.f. Dacin and Smith, 1994; Volckner and Sattler, 2006; West and Broniarczyk, 1998). Therefore, the same model is re-estimated with the alternative operationalization of variability and similar results were found. The hypothesized role of variability of past evaluations that mitigates the effect of past evaluations on new edition's evaluations persists, both in the case of consumers ($\beta = -.178$, $p < .001$) and of experts ($\beta = -.143$, $p < .001$). The results from this analysis do not support the hypothesis about the role of past sales as alternative signal for consumers when there is a high variability of evaluations ($\beta = .096$, $p = .102$). Nevertheless, this effect is still significant with regard to the experts ($\beta = .116$, $p < .05$).

3.5. Discussion

The concept of the carry over effect of product quality perception from one edition to the other has been discussed in the literature (e.g. Keller, 1993; Hennig-Thurau, Houston, Heitjans, 2009). However, past studies have not yet explored how this effect occurs. Our research is the first study that empirically explores the connection between the evaluations of one edition and the next editions. In other words, this study helps to open up the black box of the carry-over, by showing how the performance of past editions affects the evaluations of the focal edition.

The findings show that both consumers and experts are influenced by evaluations of past editions by their respective communities. This result may seem to be in contradiction with Anderson's (1973) and Oliver's (2009) papers that suggest original product's success leads to too high expectations for the sequel, eventually leading to lower satisfaction. In our study, we also find decreasing average valence of evaluation scores between original editions and the new editions, but even then, the scores are still not far apart. For example, the original *God of War* video game received a score of 94 from the critics, the sequel *God of War 2* received a score of 93, and the next sequel *God of War 3* received a score of 92. Table 3.4 presents various product editions and their evaluation scores. In the example one can see that the evaluation scores for each edition hover around the average of the series' evaluation. Studies should therefore focus more on identifying the factors that have an impact on the continuities that appear in the evaluation of series of product editions.

Table 3.4 Examples of Consumer and Expert Evaluations across Editions

| Series Name | Installment # | Consumer Evaluations | Expert Evaluations |
|--------------------|----------------------|-----------------------------|---------------------------|
| God of War | 1 | 9 | 9.4 |
| | 2 | 8.8 | 9.3 |
| | 3 | 8.6 | 9.2 |
| Battlefield | 1 | 8.3 | 8.3 |
| | 2 | 8.2 | 9.1 |
| | 3 | 7.5 | 8.5 |
| Fifa Soccer | 1('06) | 7.9 | 8 |
| | 2('07) | 7.3 | 8.2 |
| | 3('08) | 7.8 | 8.1 |
| | 4('09) | 7.9 | 8.7 |
| | 5('10) | 7.8 | 9.1 |
| | 6('11) | 8.2 | 8.9 |
| | 8('12) | 7.3 | 9 |
| | Harry Potter (movie) | 1 | 7.2 |
| | 2 | 5.8 | 7.3 |
| | 3 | 7.1 | 8.2 |
| | 4 | 7.8 | 8.1 |
| | 5 | 6.9 | 7.1 |
| | 6 | 6.8 | 7.8 |
| | 7(1st half) | 7.5 | 6.5 |
| | 7(2nd half) | 7.8 | 8.7 |

As a first step to accomplish this, this study looked at the lack of consensus among evaluators within particular communities of evaluators and presented arguments how this could operate as a signal amplifier. Past studies have mainly focused on the average of the evaluations but they largely ignored the extent of (dis-)agreement between the evaluators, and there is a call for more studies that investigate how variability in evaluations can have an effect towards the consumers (Sun, 2012). We answer this call by investigating not only the effect of lack of consensus with regard to the consumer evaluations but also with regard to the evaluations of the experts. We discuss and provide empirical evidence which shows how a lack of consensus, either in the expert or consumer communities can weaken the transferability of the communities' appreciations from one edition to the next. Considering the results that we have found, we would like to suggest that future studies should consider

looking at the level of consensus both among consumers and among expert critics when discussing new product performance, and especially if it concerns products that can be seen as part of a series.

This study also answers the call for more studies that examine the differences between experts and consumers (Zhu and Zhang, 2010). In earlier studies, evaluations by experts received most of the attention, also because these were more readily available and visible. Before the spread of broadband internet there were very limited options for consumers to publicly express their product evaluations. Apart from studies of WOM (e.g. Liu, 2006) there have been relatively few empirical papers that explore consumer reviews (Zhu and Zhang, 2010). It is also not uncommon that expert and consumer reviews are combined into one construct (e.g. Hennig-Thurau, Houston, and Heitjans, 2009). We find differences between expert and consumer evaluations and their ability to cross-influence evaluations. Interestingly, when testing a mixed model as suggested by one of the anonymous reviewers, we found that the intercept of the consumer evaluation model did not vary across series, but that it did vary for expert evaluation. This signals that on average consumers do not treat products from different series in different ways, but experts do. These results justify future research that will distinguish systematically between the two types of evaluations, each with specific antecedents and consequences. This will be possible as new editions are added to existing series, resulting in more observations per series.

Finally, as a minor finding, this study shows that experts may not be as objective as they are sometimes assumed to be. Future research should take into account that experts may be affected by how they see the consumers reacting to the product or franchise, by observing consumer reviews or sales. The desire to maximize profit by the media that employ them can force experts to heed the market (Shoemaker and Vos, 2009), but more independent experts may also fear to lose relevance to their readers if their opinions drift too far away from the

average consumer's. Research on reviews usually assumes a unidirectional causality, i.e. from the expert reviews on the behavior of the consumers, and overlooked the possibility that the evaluations of experts should be considered to be partly endogenous. The approach to analyze evaluations across editions gives an opportunity to explore how past market performance can influence future expert judgments, which suggests a two way relationship between the market and the expert.

The results of this study suggest a number of managerial implications. The exploration of the determinants of evaluations serves to better understand and predict whether the popularity of a franchise warrants another edition. Another important implication is that producers of goods that are produced in series should not take mediocre expert reviews too much into account if consumer reviews and sales are good, because there is a good probability that the experts will revise their opinions in respect of later editions without the producers making additional efforts or costs to satisfy their preferences. However, if the earlier versions have not been too popular among consumers it does make more sense to also attempt to turn the expert reviewers around.

This paper has several limitations. First, the dataset used in the analysis is based on numerical evaluation scores submitted by consumers and experts. Future studies may also employ textual data mining tools that can generate evaluation scores based on the content of the evaluations. Such tools can also generate more robust measures of community consensus based on the dissimilarities of the wording and not just the variance of the scores. However, such an analysis would require a very large database and a powerful computation algorithm because the number of texts to be analyzed can reach hundreds, if not thousands, from each type of evaluator, for each edition. This study is based on a dataset from the video games industry, using inputs from aggregator websites. These websites are less readily available outside of the creative industries, but, newer data-mining tools will allow much wider ranging

explorations of un-aggregated evaluations on the internet, that are on individual blogs, forums, social network sites, or other online type of (online) communities. Future research will be able to capitalize on these developments and identify patterns, as well as study in depth the degree of consensus within and between different communities of evaluators. Additionally, the evaluations in the dataset used in this study only have a single dimension. It is not uncommon that consumers and experts provide their evaluations in respect to several dimensions, for example, product enjoyment, ease of use, utility, artistic value. Future research that would include multiple dimensions along which evaluations are scored could further explore the extent to which scores along particular dimensions of past editions have an effect on scores along the same or different dimensions in the evaluations of the new edition. Further studies that extend the approach followed by this study to other industries may also have to include more or other industry-specific signals that may affect the carry-over from one edition to the next. For example, in analyzing signals in the movie industry, it is important to control for star power (Ravid, 1999). This study does control for possible changes of developer between the editions. This study finds this factor is insignificant, but this may be due to the relatively low importance of software developers, programmers, or project directors as "stars" whose name carries enough reputational value to influence experts and consumers' judgment in the video games industry, future studies that extend our model but based on other industry, e.g. the movie industry, may find this factor significant. The study is limited to new editions, without taking into account possible spinoffs of the franchise. For example, after the success of the book series, the Harry Potter franchise created millions of dollars revenue from movie productions and video games. Future studies may investigate how expert and consumers evaluations can be transferred not only to the next edition of the series, but also to extensions in other media.