



Thank you for downloading this document from the RMIT Research Repository.

The RMIT Research Repository is an open access database showcasing the research outputs of RMIT University researchers.

RMIT Research Repository: <http://researchbank.rmit.edu.au/>

Citation:

See this record in the RMIT Research Repository at:

Version:

Copyright Statement: ©

Link to Published Version:

PLEASE DO NOT REMOVE THIS PAGE

**HISTORY MATTERS: THE IMPACT OF REVIEWS AND SALES OF
EARLIER VERSIONS OF A PRODUCT ON CONSUMER AND
EXPERT REVIEWS OF NEW EDITIONS**

Frederik B.I. Situmeang

PhD Candidate in Marketing and Strategy
Amsterdam Business School
University of Amsterdam
Plantage Muidergracht 12
1018TV Amsterdam
f.b.situmeang@uva.nl
+31-(0)20 5258149
The Netherlands

Mark A.A.M. Leenders

Professor of Marketing
Royal Melbourne Institute of Technology
124 La Trobe Street
3000, Melbourne
mark.leenders@rmit.ac.au
+61-(0)3 9925 1582
Victoria, Australia

Nachoem M. Wijnberg

Professor of Cultural Entrepreneurship and Management
Amsterdam Business School
University of Amsterdam
Plantage Muidergracht 12
1018TV Amsterdam
n.m.wijnberg@uva.nl
+31-(0)20 5256106
The Netherlands

This paper has been accepted and is forthcoming in the European Management Journal

HISTORY MATTERS: THE IMPACT OF REVIEWS AND SALES OF EARLIER VERSIONS OF A PRODUCT ON CONSUMER AND EXPERT REVIEWS OF NEW EDITIONS

ABSTRACT

Product reviews are assumed to be based on the observable characteristics of the underlying product. However, in the case of new editions in a product series, the determinants may include signals that originate from the reviews and the sales of editions that precede the focal product edition. Our analysis of 577 video games released in a series between 2000 and 2009 indicates that the reviews of earlier versions carry over to the reviews of the sequel by the same type of reviewer. We also find that expert reviews are influenced by the average review of previous editions by consumers and the average sales of previous editions 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 reviewers of a particular type, weakens the impact of average past reviews, whilst it magnifies the impact of the sales of earlier versions.

Keywords: Sequels, product reviews, expert critics, consumers, variability

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 a 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 impressions connected to the earlier editions are transferred to the next edition (Keller 2003; Erdem, 1998; Situmeang, Leenders, and Wijnberg, forthcoming; 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 edition generates high expectations that can easily lead to a lower level of satisfaction (Anderson, 1973; Oliver, 2009) and lower sales than that of the original editions (Basuroy and Chatterjee, 2008). Thus, there are explanations why a sequel would do better and reasons why it would do worse. The managerial assumption is that sequels, preserving the same core attributes in the same series, experience similar or better sales than the previous version. This paper, however, does not analyze the

factors that determine sales performance but the factors that lead to changes in the reviews of sequel products, among which the past sales performance will figure as a predictor.

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 reviews 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 reviews are usually treated as independent determinants of sales performance. At the same time, it is usually assumed that product reviews (or evaluations) 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 reviews 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 reviews of the new editions will not differ much from reviews 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 study looks more closely at the question whether the reviews of products in a series are stable over the series of editions and, if they are not, what the main predictors of the variation are.

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 reviews of both consumers and experts (and their sales performance), on the expert and consumer reviews of the new edition.

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

Our study contrasts with earlier studies (e.g. Keller, 1993; Hennig-Thurau, Houston and Heitjans, 2009) in which consumer reviews and expert critic reviews are 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 reviews do have similar effects on sales, the determinants of the reviews 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 reviews 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 reviews are influential factors for consumer buying behavior, such as the automotive industry, cellphones, or laptops among many others.

2. Theory and Hypotheses

2.1. Product reviews as signals

Product reviews, 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 reviews 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) reviews and almost half of them (49%) read the reviews 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 reviews (Deloitte, 2007).

The prominence of online reviews has led many scholars to explore how reviews affect sales, with initial studies focusing on the impact of expert reviews (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 reviews 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 reviews and expert reviews 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 reviews on sales, there are fewer studies that investigate the determinants of the product reviews themselves. Rather it is assumed that reviews are simply reflections of how well the characteristics of the product fit quality standards and the preferences of the reviewer. It is assumed that the signals that are taken into account to arrive at the review are those that 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 is 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 review 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 reviews of earlier versions, and 2) the sales performance of earlier versions.

With regard to the effects of earlier product reviews, the default expectation would be continuity; such that if the previous edition received on average a positive review score from a community of reviewers, the focal edition will also be evaluated positively by that same community. However, the average review tells only part of the story. Another signal that is important is the variability of reviews within each community of reviewers - experts or consumers - as this reflects the degree of consensus among reviewers within their respective

community. This can potentially influence the effects of past reviews and past sales. The role of variability as an amplifier of peripheral signals on the reviews of the focal edition will be discussed and tested more extensively in the following sections.

2.2. The relationship between reviews of past editions and the reviews of new editions

Products with the same brand name are often closely linked to each other, and 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 reviews are also *signals* of consumers and experts appreciations of *the series*, thus the review affects not only the sales of the particular editions that the reviews 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

reviews by consumers of earlier editions in the same series not only signal attractiveness regarding the product to consumers, but potentially also the attractiveness of the franchise or series as a whole. Considering this aspect of consumer reviews, as indicators of the quality as perceived by consumers of the series, one can expect some form of continuity in the average response of consumers regarding the subsequent editions of the series. For this effect to operate, one does not have to assume that consumers have extensive knowledge of the consumer reviews of the previous editions. It suffices that these reviews are at least partly influenced by characteristics that are attributed to the series as a whole. This suggests:

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

The general arguments presented above with respect to consumer reviews, is likely also applicable to the relationship of expert reviews of past editions with expert reviews of the new edition. New editions or sequels are generally similar to the original product in terms of core product attributes and 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. Expert reviewers might be more wary of publicly changing their opinions about particular franchises, since their professional reputation will be linked to their reviews and changing their mind might harm their credibility. Consequently, expert opinions of the new edition would be very similar to the reviews 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 reviews will remain even more constant than consumers. However, there is also an argument suggesting an effect in the opposite direction. Professional critics are more likely to notice a lack of creativity and innovation (e.g. Hudson, 2011; Kondoloy, 2012), which suggests that maintaining similar product attributes would lead to a downward trend of expert review 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 reviews of past editions in the series have a positive relationship with expert reviews of the new edition of the product.

2.3. The relationship between sales performance of past editions and the reviews of the 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). Past studies have shown that market performance of the previous editions in a series has a significant and positive effect towards the market performance of the new edition (Hennig-Thurau, Houston and Heitjans, 2009; Situmeang, Leenders and Wijnberg, forthcoming). One can also adopt the concept of signal theory to explain the previous relationship. Considering that the main purpose of product review as a signal is to reduce the uncertainty about the product by signaling to the consumers what can be expected from it, one can consider past sales as a periphery signal

because it can serve as a reflection, to some extent, on how well the series has been popular or 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 reviews 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 reviews of the new edition of the product.

It may be argued that past sales are less likely to influence critics. Consumer reviews 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* reviewers, 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 or 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 reviews 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 reviews of the new edition of the product.

2.4. Consensus in the community of reviewers as a signal-amplifier

In most social groups reviews 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 reviews 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 reviews (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 reviews of books and the standard deviation of these reviews 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 reviews, not of demand. Moreover, we will look at variability as a moderator of the effects of the other signals, such as average past reviews.

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 reviewers as a group to assess the qualities of these past editions. Although expert reviewers may be more likely to pay explicit attention, consumers do as well observe and notice the differences and the distributions of product reviews across time (Chevalier and Mayzlin, 2006; Sun, 2012). Given that, as we argued above, average reviews 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 reviews 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 reviewers 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 between past consumer reviews with consumer reviews of the new edition of the product is moderated by the degree of consensus among consumer reviewers 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 reviews of past editions become less convincing as a signal of the qualities of the whole series when consensus among expert reviewers 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 reviews of past editions. This suggests:

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

As has previously noted, the main function of reviews 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 reviewers

of a particular type makes the average reviews of past editions less clear or trustworthy signals to the reviewers, of the same type, of new editions.

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

H4a: The relationship of past sales with consumer reviews of the new edition of the product is moderated by the degree of consensus among consumer reviewers 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 reviewers 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 reviews of past editions, may be greater with regard to the experts than to the consumers. If expert reviewers 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 reviews of the new edition of the product is moderated by the degree of consensus among expert reviewers of the past editions. The relationship is more positive if there is a lower degree of consensus.

3. Material and Methods

3.1. The sample

To test the relationships of the peripheral signals with the reviews of new editions, we benefit from online data repository that provides extensive performance records of creative products in terms of sales and reviews (vgchartz.com for the sales and metacritic.com for the reviews). These online databases have also been used in other 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.

3.2. The operationalization of the variables

Expert Reviews and *Consumer Reviews* 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 review, the average of past reviews in the series is measured by averaging relevant average review scores of past editions in the series. For example, the average consumer review of past editions in the series is measured by averaging the average consumer review scores per edition in the series before the focal product edition.

The *Lack of Consensus* denotes the variability of reviews given by the reviewers 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 review scores was used 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 review.

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

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 reviews, the average past unit sales is measured by calculating the average of worldwide unit 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 an approach based on units 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 reviews 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.

4. Results

4.1. Correlation Matrix and Descriptive Analysis

Table 1 presents the correlation matrix and the descriptive statistics of the variables used in this research. On average sequels in the sample received positive reviews from both experts and consumers, both review scores average above 7 out of 10. On the contrary, it seems that expert reviews ($\mu = 4.96$) display a larger variability than consumer reviews ($\mu = 3.75$), 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 review scores of the new editions are highly correlated with the respective community past reviews 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 reviews ($r = .107$, $p < .05$) and unexpectedly an even stronger relationship with the expert reviews ($r = .226$, $p < .001$). Most importantly, there is no correlation above $r = .5$ between the independent variables; therefore, the independence of the constructs is verified.

[Table 1 about here]

4.2. Pre-Test: Endogeneity and Selection Bias

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. Indeed, the Hausman test on our main models suggest endogeneity (experts review model $p = .022$; consumer review model $p = .045$). Ho, Dhar and Weinberg (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. Ho, Dhar and Weinberg (2009) first predict the probability of airing a trailer using instrumental variables (not correlated with the outcome variable) and the estimated probability from this stage is used as a predictor in the main model. In our case, the goal is to include a parameter for the probability to develop a sequel in the first place. Our main instrumental variables in the first stage model are the quantity of competing products (*COMP*), mode of play: Multiplayer (*MULTI*) or Single player (the default variable), ESRB age restriction: “*TEEN*”, “*MATURE*” or “Everyone” (the default variable), and genre classifications: First Person Shooter (*FPS*), *RACING*, Role Playing Games (*RPG*), and Sports. Competition (*COMP*) is measured by the number of other products (i.e. not created by the same company that developed the focal product) in the same year which occupied the same genre as the focal product. The function is mathematically written as follows:

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

This model 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 review and expert review of the sequel (cf. Greene, 2003: 334; Ho, Dhar, and Weinberg, 2009).

To confirm that we have tackled the issue of endogeneity (and the issue of omitted or

confounding variables), we conducted an additional correlation test between our independent variables and the error terms of our main models. This confirms that there are no significant correlations between the error terms and our independent variable. In addition to this, we also conducted a test of over-identification to examine the appropriateness of our instruments. We do not reject the null hypothesis, implying that our sets of instruments are largely orthogonal (i.e., not correlated) with the error term, therefore, confirming the results from the correlation test (expert reviews model test = 5.381, $p = .614$; consumer reviews model test = 9.759, $p = .203$).

4.3. Test of the 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^{1,2}:

$$\begin{aligned} ConsRev_i = & \alpha + \beta_1 AvgConsRevPE_i + \beta_2 AvgExpertRevPE_i + \beta_3 VarConsRevPE_i \\ & + \beta_4 AvgSalesPE_i + \beta_5 AvgConsRevPE * VarConsRevPE_i \\ & + \beta_6 AvgSalesPE * VarConsRevPE_i + \beta_7 AvgExpertRevPE \\ & * VarConsRevPE_i + \beta_8 ChangeAge_i + \beta_9 ChangeGenre_i \\ & + \beta_{10} ChangeDev_i + \beta_{11} NumberEdition_i + \beta_{12} Nintendo_i + \beta_{13} Sony_i \\ & + \beta_{14} ExpertRev_i + \beta_{15} \hat{p}(Sequel)_i + \varepsilon_i \end{aligned}$$

$$\begin{aligned} ExpertRev_i = & \gamma + \vartheta_1 AvgConsRevPE_i + \vartheta_2 AvgExpertRevPE_i + \vartheta_3 VarExpertRevPE_i \\ & + \vartheta_4 AvgSalesPE_i + \vartheta_5 AvgExpertRevPE * VarExpertRevPE_i \\ & + \vartheta_6 AvgSalesPE * VarExpertRevPE_i + \vartheta_7 AvgConsRevPE \\ & * VarExpertRevPE_i + \vartheta_8 ChangeAge_i + \vartheta_9 ChangeGenre_i \\ & + \vartheta_{10} ChangeDev_i + \vartheta_{11} NumberEdition_i + \vartheta_{12} Nintendo_i + \vartheta_{13} Sony_i \\ & + \vartheta_{14} \hat{p}(Sequel)_i + \omega_i \end{aligned}$$

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

¹ c.f. Byrne, 2000: 11

² 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 Beta-Coeff. = .140 (.013) in the mixed model, while it was Beta-Coeff. = .122(.010) in our original AMOS model.

consumer review and .297 for expert review. A summary of results of the analysis is reported in Table 2.

[Table 2 about here]

The results of the analysis of our model suggest that consumer reviews of the new editions are positively related to the average consumer reviews across past editions ($\beta = .219$, $p < .001$), average expert review 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 review is significantly lowered ($\beta = -.127$, $p < .01$) by the lack of consensus between consumers. This relationship is visualized in Figure 1. 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 2. These results offer support for H1a, H2a, H3a, and H4a.

[Figure 1 about here]

[Figure 2 about here]

The results also provide strong evidence that expert reviews of the new editions are highly affected by average expert reviews 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 reviews across past editions with expert reviews of the new edition ($\beta = .186$, $p < .001$). The positive effect of past expert reviews across editions is significantly lowered ($\beta = -.144$, $p < .01$) by the lack of consensus between the experts (visualized in Figure 3). 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 about here]

[Figure 4 about here]

Consistent with the results of earlier studies on sequels, we found a negative effect of the number of editions in the series on product review (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 reviews of sequels of Xbox and of Nintendo games by experts ($\beta = .003$, $p > .1$) or consumers ($\beta = .011$, $p > .1$). Experts reviews are not significantly different for PlayStation games ($\beta = .052$, $p > .1$), though consumer reviews 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 reviews, and this effect is only limited to expert reviews ($\beta = -.113$, $p < .01$). The significance level of the other control variables is higher than .10, suggesting that their impact on the expert and consumer reviews is negligible.

4.4. Post analysis and 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 reviews in the sample used in this study are also skewed to the same direction (skewness = -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 reviews. 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 and they are very similar to the initial findings. The 3SLS Multiple R-squares are .230 and .254 respectively for the consumers and experts reviews 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 review scores, and the lack of consensus in past reviews 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 about here]

Next, instead of using the review 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 reviews of new editions ($\beta = .064$, $p = .214$) or on the number of expert reviews ($\beta = .062$, $p = .199$). It seems likely that, because of strong relationship between sales and number of reviewers, the effects of past sales are absorbed in the effects of the number of past reviewers. 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 review scores within the editions, the variability of the average reviews across editions is used to measure the lack of consensus, please refer to the following formula:

$$Variability_Alternative_{i,j} = Var[Mean(Review)_{0\dots i-1,j}]$$

Arguably, the two are conceptually different factors; the first measures the degree of the lack of consensus about a single object of review in a community of reviewers, 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 reviews that mitigates the effect of past reviews on new edition's reviews 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 reviews ($\beta = .096$, $p = .102$). Nevertheless, this effect is still significant with regard to the experts ($\beta = .116$, $p < .05$).

5. Discussion and Conclusion

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). Our research is the first study that theorizes on and empirically explores the connection between consumer and expert reviews in a series of editions and the sales of previous editions. In other words, this study helps to open up the black box of key carry-over mechanisms between previous editions and the latest edition in the series. In addition, the role of variability and consensus is studied in more detail.

The findings show that both consumers and experts are influenced by reviews of past editions by their respective communities. This result may seem to be counterintuitive if compared to 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 review 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 4 presents various product editions and their review scores. In the example one can see that the review scores for each edition hover around the average of the series' review. Studies should therefore focus on identifying the factors that have an impact on the continuities that appear in the review of series of product editions.

[Table 4 about here]

As a first step to accomplish this, this study looked at the lack of consensus among reviewers within particular communities of reviewers and presented arguments how this could operate as a signal amplifier. Past studies have mainly focused on the average of the reviews but they largely ignored the extent of (dis-)agreement between the reviewers, and

there is a call for more studies that investigate how variability in reviews 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 reviews but also with regard to the reviews by 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, reviews 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 reviews. 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 reviews and their ability to cross-influence reviews. Interestingly, when testing a mixed model as suggested by one of the anonymous reviewers, we found that the intercept of the consumer review model did not vary across series, but that it did vary for expert review. 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 reviews, 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 reviews of experts should be considered to be partly endogenous. The approach to analyze reviews 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 reviews 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 review scores submitted by consumers and experts. Future studies may also employ textual data mining tools that can generate review scores based on the content of the reviews. 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 reviewer, 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 unaggregated reviews 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 reviewers. Additionally, the reviews in the dataset used in this study only have a single dimension. It is not uncommon that consumers and experts provide their reviews in respect to several dimensions, for example, product enjoyment, ease of use, utility, artistic value. Future research that would include multiple dimensions along which reviews 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 reviews 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. It is very common for companies to capitalize on the success of their products by creating product extensions in various media other than the original product's. 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 reviews can be transferred not only to the next edition of the series, but also to extensions in other media.

References

- Aaker, D.A., & Keller, K.L. (1990). Consumer evaluations of brand extensions. *Journal of Marketing*, 54(1), 27–41.
- Adams, E., & Rollings, A. (2003). Andrew Rollings and Ernest Adams on game design. New Riders Publishing.
- Allen, M.P., & Lincoln, A.E. (2004). Critical discourse and the cultural consecration of American films. *Social Forces*, 82(3), 871-894.
- Allen, M.P., & Parsons, N.L. (2006). The institutionalization of fame: Achievement, recognition, and cultural consecration in baseball. *American Sociological Review*, 71(5), 808-825.
- Arbuckle, J. L. (2009). *AMOS 18 users' guide*. Chicago: SmallWaters.
- Basuroy, S., Chatterjee, S., Ravid, A. (2003). How critical are critical reviews? The box office effects of film critics, star power, and budgets. *Journal of Marketing*, 67(4), 103-117.
- Basuroy, S., Chatterjee, S. (2008). Fast and frequent: Investigating box office revenues of motion picture sequels. *Journal of Business Research*, 61(7), 798-803.
- Beam, R.A. (2000). Content differences between daily newspapers with strong and weak market orientations. *Journalism and Mass Communication Quarterly*, 80(2), 368-90.

- Bhansing, P.V., Leeenders, M.A.A.M., & Wijnberg, N.M. (2012). Performance effects of cognitive heterogeneity in dual leadership structures in the arts: The role of selection system orientations. *European Management Journal*, 30(6), 523-534.
- Boatwright, P., Basuroy, S., Kamakura, W. (2007). Reviewing the reviewers: The impact of individual film critics on box office performance. *Quantitative Marketing and Economics*, 5(4), 401-425.
- Boor, M. (1990). Reliability of ratings of movies by professional movie critics. *Psychological Reports*, 67(1), 243-257.
- Byrne, B.M. (2000). *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming*. Lawrence Erlbaum Associates. Mahwah, NJ.
- Cameron, S. (1995). On the role of critics in the culture industry. *Journal of Cultural Economics*, 19(4), 321–331.
- Caves, R.E. (2000), *Creative industries*. Cambridge, MA: Harvard University Press.
- Chen, Y., Liu, Y., & Zhang, J. (2012). When Do Third-party Product Reviews Affect Firm Value and What Can Firms Do? The Case of Media Critics and Professional Movie Reviews. *Journal of Marketing*, 76(2), 116-134.
- Chevalier, J., & Mayzlin, D. (2006). The effect of word of mouth online: Online book reviews. *Journal of Marketing Research*. 43(3), 345–354.
- Choi, Y.B., (1993). *Paradigms and Conventions: Uncertainty, decision making, and entrepreneurship*. Michigan: The University of Michigan Press.
- Dacin, P.A., & Smith, D.C. (1994). The effect of brand portfolio characteristics on consumer evaluations of brand extensions. *Journal of Marketing Research*, 31(2), 229-42.
- Das, S.R., & Chen, M.Y. (2007). Yahoo! For amazon: Sentiment extraction from small talk on the web. *Management Science*, 53(9), 1375-1388.
- Debenedetti, S. (2006). The Role of Media Critics in the Cultural Industries. *International Journal of Arts Management*, 8(3), 30-42.
- Deloitte (2007). Most consumers read and rely on online reviews; companies must adjust. (accessed 5 March 2012) [available at www.marketingcharts.com/interactive/most-consumers-read-and-rely-on-online-reviews-companies-must-adjust-2234/]
- Duan, W., Gu, B., & Whinston, A.B. (2008). The dynamics of online word-of-mouth and product sales: an empirical investigation of the movie industry. *Journal of Retailing*, 84 (2), 233–42.

- Dwyer, F. R., & Welsh, M.A. (1985). Environmental relationships of the internal political economy of marketing channels. *Journal of Marketing Research*, 22(4), 397-414.
- Eliashberg, J., & Shugan, S.M. (1997). Film critics: influencers or predictors? *Journal of Marketing*, 61(April), 68-78.
- Erdem, T. (1998), An Empirical Analysis of Umbrella Branding, *Journal of Marketing Research*, 35 (August), 339–51.
- Gemser, G., Oostrum, M. van, Leenders, M.A.A.M. (2007). The impact of film reviews on the box office performance of art house versus mainstream motion pictures. *Journal of Cultural Economics*, 31(1), 43-63.
- Graffin S, and Ward, A. (2010). Certifications and reputation: Determining the standard of desirability amidst uncertainty. *Organization Science*, 21(2), 331-346.
- Greene, W.H. (2003). *Econometric analysis*, 5th ed.: Prentice Hall.
- Hennig-Thurau, T., & Houston, M.B., & Heitjans, T. (2009). Conceptualizing and measuring the monetary value of brand extensions: The Case of Motion Pictures. *Journal of Marketing*, 73(6), 167-183.
- Hill, C.W.L. (1990). Cooperation, opportunism, and the invisible hand: implications for transaction cost theory. *The Academy of Management Review*, 15(3), 500-513.
- Ho, J.Y.C. , Dhar, T., & Weinberg, C.B. (2009). Playoff payoff: Super Bowl advertising for movies. *International Journal of Research in Marketing*, 26(3), 168-179.
- Hoyer, W.D. (1984). An examination of consumer decision making for a common repeat purchase product. *Journal of Consumer Research*, 11(3), 822-829.
- Hsu, G. (2006). Evaluative schemas and the attention of critics in the US film industry. *Industrial and Corporate Change*. 15(3), 467-496.
- Hu, N., Pavlou, P.A., & Zhang, J. (2006). Can online reviews reveal a product's true quality?: empirical findings and analytical modeling of Online word-of-mouth communication. *EC '06 Proceedings of the 7th ACM conference on Electronic commerce*.
- Hudson, A. (2011). Have computer games become boring? (accessed 5 March 2012) [available at http://news.bbc.co.uk/2/hi/programmes/click_online/9538690.stm]
- Karniouchina, E.V. (2011). Impact of star and movie buzz on motion picture distribution and box office revenue. *International Journal of Research in Marketing*, 28(1), 62-74.
- Keller, K.L. (1993). Conceptualizing, Measuring, and Managing Customer-Based Brand Equity, *Journal of Marketing*, 57(1), 1–22.

- Keller, K.L. (2003). *Strategic brand management: Building, measuring, and managing brand equity* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall.
- Kirmani, A. & Rao, A.R. (2000). No pain, no gain: a critical review of the literature on signaling unobservable product quality. *The Journal of Marketing*, 64(2), 66-79.
- Kondoljy, A.L. (2012). How The Video Game Industry Is Becoming Like Hollywood. (accessed 5 March 2012) [available at <http://www.cheatcc.com/extra/howthevideogameindustryisbecominglikehollywood.htm>]
- Lightspeed Research (2011). Consumers rely on online reviews and price comparisons to make purchase decisions. (accessed 5 March 2012) [available at <http://www.lightspeedresearch.com/press-releases/consumers-rely-on-online-reviews-and-price-comparisons-to-make-purchase-decisions/>]
- Moon, S., Bergey, P., & Iacobucci, D. (2010). Dynamic Effects Among Movie Ratings, Movie Revenues, and Viewer Satisfaction. *Journal of Marketing*, 74(1), 108-121.
- Nevitt, J. & Hancock, G.R. (2001). Performance of bootstrapping approaches to model test statistics and parameter standard error estimation in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 8(3), 353-377.
- Neuwirth, K., Frederick, E., & Mayo, C. (2007). The spiral of silence and fear of isolation. *Journal of Communication*, 57(3), 450-468.
- Oliver, R. L. (2009), *Satisfaction: A Behavioral Perspective on the Consumer*. New York: McGraw-Hill.
- Oshagan, H. (1996). Reference group influence on opinion expression. *International Journal of Public Opinion Research*, 8(4), 335-354.
- Perks, H., & Halliday, S.V. (2003). Sources, signs and signalling for fast trust creation in organisational relationships. *European Management Journal*, 21(3), 338-350.
- Phillips, C. (2011). Sequels dominate fall video games releases. (accessed 5 March 2012), [available at <http://www.timesfreepress.com/news/2011/sep/02/0902-e1-game-developers-unleashing-a-wave-of/>]
- Plucker, J.A., Kaufman, J.C., Temple, J.S., & Qian, M. (2009). Do experts and novices evaluate movies the same way? *Psychology and Marketing*, 26(5), 470-478.
- Reddy, S., Swaminathan, V., & Motley, C. (1998). Examining the Determinants of Broadway Show Success, *Journal of Marketing Research*, Vol. 35, August, pp. 370–383.

- Rees, C.J. van (1987). How reviewers reach consensus on the value of literary works. *Poetics*, 16(3-4), 275-294.
- Rosen, D. (2011). The other side: Creative bankruptcy. (accessed 5 March 2011), [available at <http://www.callitlikeiseeit.com/230>]
- Rothschild, M.L., & Gaidis, W.C. (1981). Behavioral learning theory: Its relevance to marketing and promotions. *The Journal of Marketing*, 45(2), 70-78.
- Shoemaker, P.J., & Vos, T.P. (2009). *Gatekeeping Theory*. Oxon: Routledge.
- Situmeang, F.B.I., Leenders, M.A.A.M., & Wijnberg, N.M. (forthcoming). The good, the bad and the variable: how evaluations of past editions influence the success of sequels. *European Journal of Marketing*.
- Sood, S., & Dreze, X. (2006). Brand extensions of experiential goods: Movie sequel evaluation. *Journal of Consumer Research*, 33(3), 352-360.
- Spence, M. (1973). Job Market Signaling. *The Quarterly Journal of Economics*, 87(3), 355-374.
- Sun, M. (2012). How does variance of product ratings matter? *Management Science*, 58(4), 696-707.
- Sunde, L., & Brodie, R.J. (1993). Consumer evaluations of brand extensions: further empirical results. *International Journal of Research in Marketing*, 10(1), 47-53.
- Tschang, F.T., & Szczypula, J. (2006). Idea creation, constructivism and evolution as key characteristics in the videogame artifact design process. *European Management Journal*, 24(4), 270-287.
- Volckner, F., & Sattler, H. (2006). Drivers of brand extension success. *Journal of Marketing*, 70(2), 18-34.
- West, P.M. & Broniarczyk, S.M. (1998). Integrating multiple opinions: The role of aspiration level on consumer response to critic consensus. *Journal of Consumer Research*, 25(1), 38-51.
- Wijnberg, N.M., & Gemser, G. (2000). Adding value to innovation: Impressionism and the transformation of the selection system in visual arts. *Organization Science*, 11(3), 323-329.
- Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing*, 74(2), 133-148.

Table 1. Descriptive Statistics and Correlation Matrix

	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12
1	Consumer Reviews of the Sequel	7.814	1.491											
2	Expert Reviews of the Sequel	7.437	1.257	.353***										
3	Average Consumer Reviews of Past Editions (PE)	8.021	1.086	.269***	.085									
4	Variability of Consumer Reviews of PE	3.746	4.145	.010	-.038	.057								
5	Average Quantity of Consumer Reviewers of PE	44.711	135.624	.127***	-.048	.121***	-.020							
6	Average Expert Reviews of PE	7.198	2.150	.103*	.319***	.253***	-.015	.163***						
7	Variability of Expert Reviews of PE	4.963	6.522	-.072	-.140***	-.152***	-.008	-.104**	-.224***					
8	Average Quantity of Expert Reviewers 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

*** significant at < .001

** significant at <.01

* significant at <.05

Table 2. Standardized Path Coefficients

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

Sig. levels are two tailed

Table 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.
<i>Simple Effects</i>				
Avg. Cons. Rev. Past Editions (PE) → Cons. Rev. Sequel (S)	.357	<.001	.181	<.001
Avg. Expert Rev. PE → Expert Rev. S	.381	<.001	.261	<.001
Avg. Expert Rev. PE → Cons. Rev. S	.071	.218	.008	.883
Avg. Cons. Rev. PE → Expert Rev. S	.138	.008	.071	.164
Avg. Sales PE → Cons. Rev. S	.106	.052	.136	.004
Avg. Sales PE → Expert Rev. S	.119	.001	.182	.001
Var. Cons. Rev. PE → Cons. Rev. S	-.005	.489	-.022	.710
Var. Expert Rev. PE → Expert Rev. S	-.096	.002	-.432	<.001
<i>Interaction Effects</i>				
Avg. Cons. Rev. PE * Var. Cons. Rev. PE → Cons. Rev. S	-.164	.006	-.109	.036
Avg. Expert Rev. PE * Var. Expert Rev. PE → Expert Rev. S	-.334	<.001	-.165	.017
Avg. Expert Rev. PE * Var. Cons. Rev. PE → Cons. Rev. S	.010	.459	.039	.471
Avg. Sales PE * Var. Cons. Rev. PE → Cons. Rev. S	.005	.891	.014	.794
Avg. Cons. Rev. PE * Var. Expert Rev. PE → Expert Rev. S	.027	.702	.018	.763
Avg. Sales PE * Var. Expert Rev. PE → Expert Rev. S	.098	.017	.053	.244

Sig. levels are two tailed

Table 4. Examples of Consumer and Expert Reviews across Editions

Series Name	Installment #	Consumer Reviews	Expert Reviews
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

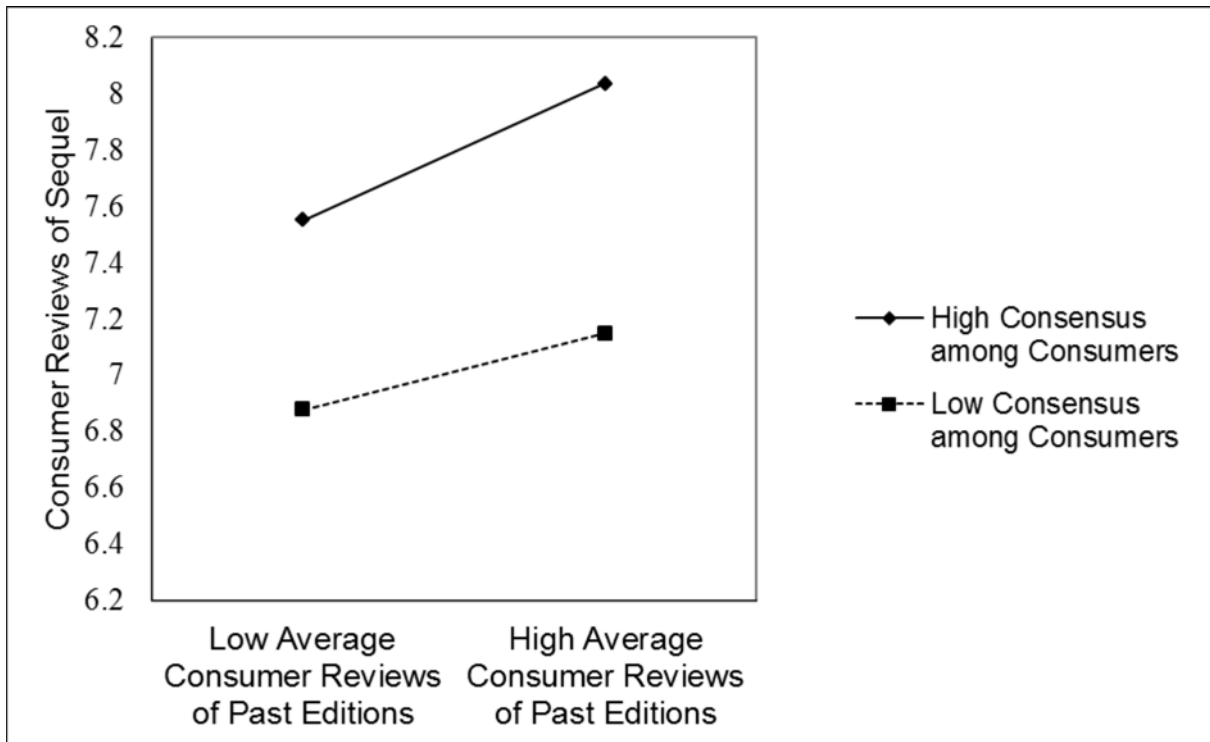


Figure 1. The effect of consumer reviews of past editions on consumer reviews of the new edition (sequel) with low or high variability of consumer reviews

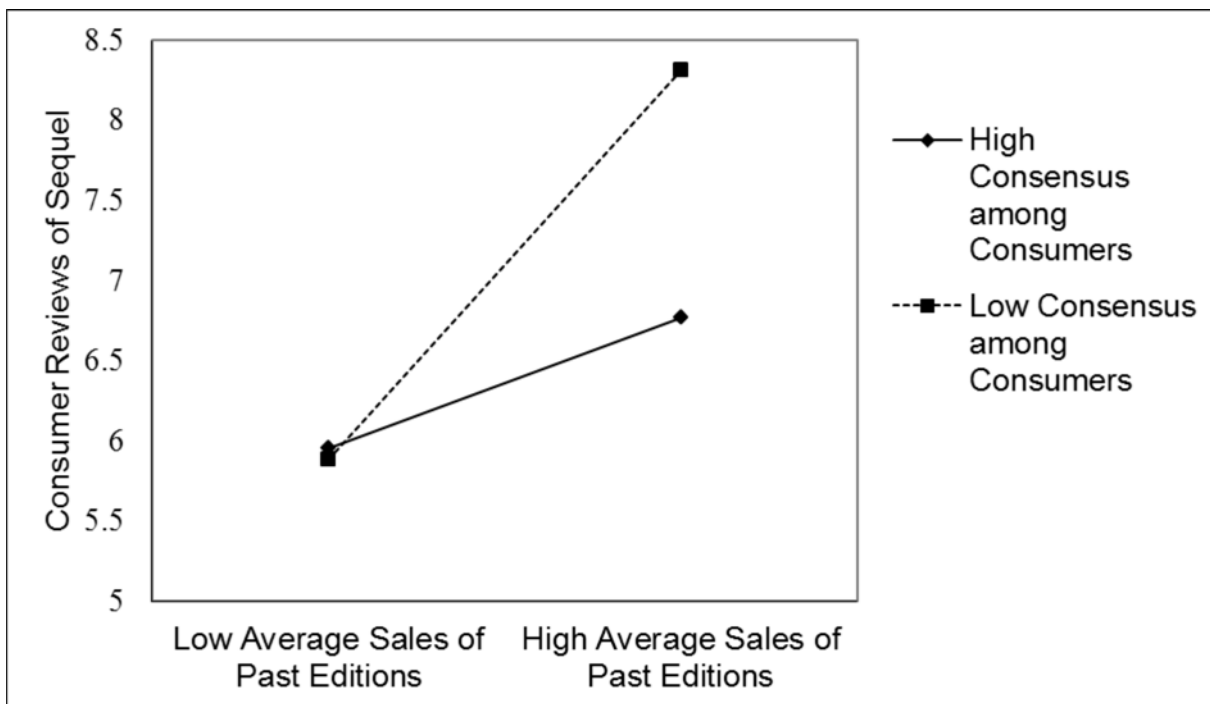


Figure 2. The effect of sales of past editions on consumer reviews of the new edition (sequel) with low or high variability of consumer reviews

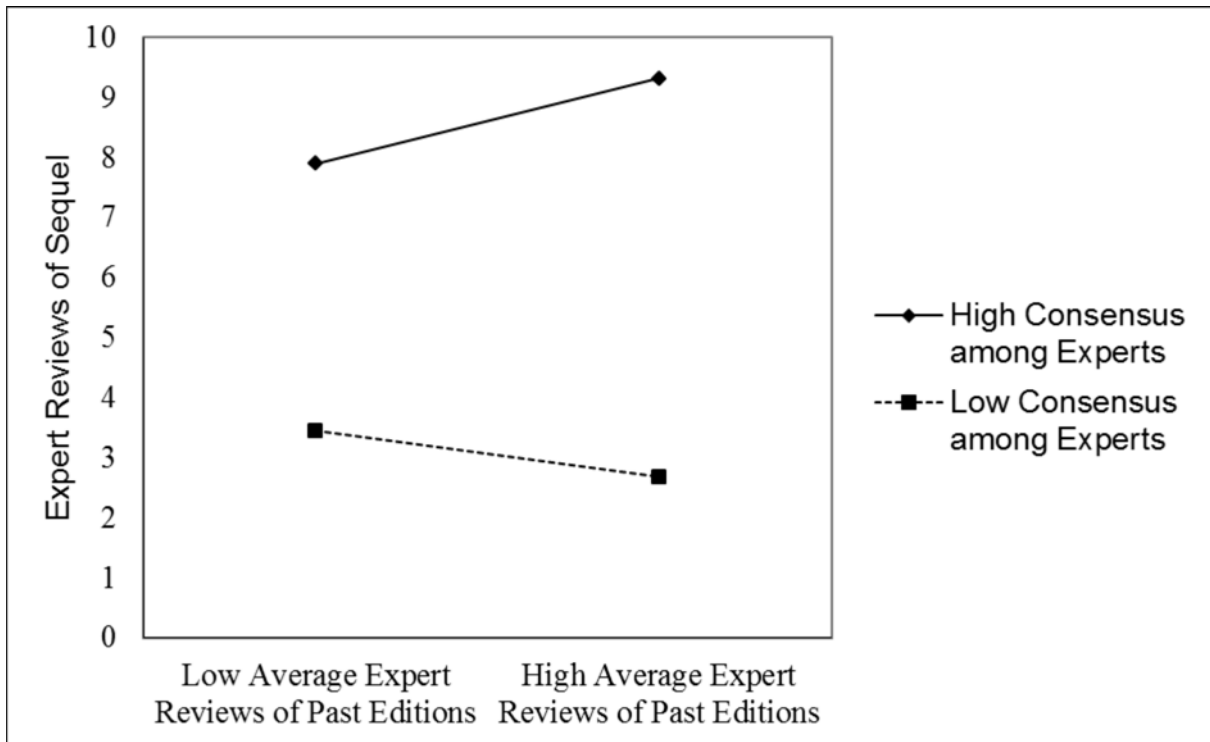


Figure 3. The effect of expert reviews of past editions on expert reviews of the new edition (sequel) with low or high variability of expert reviews

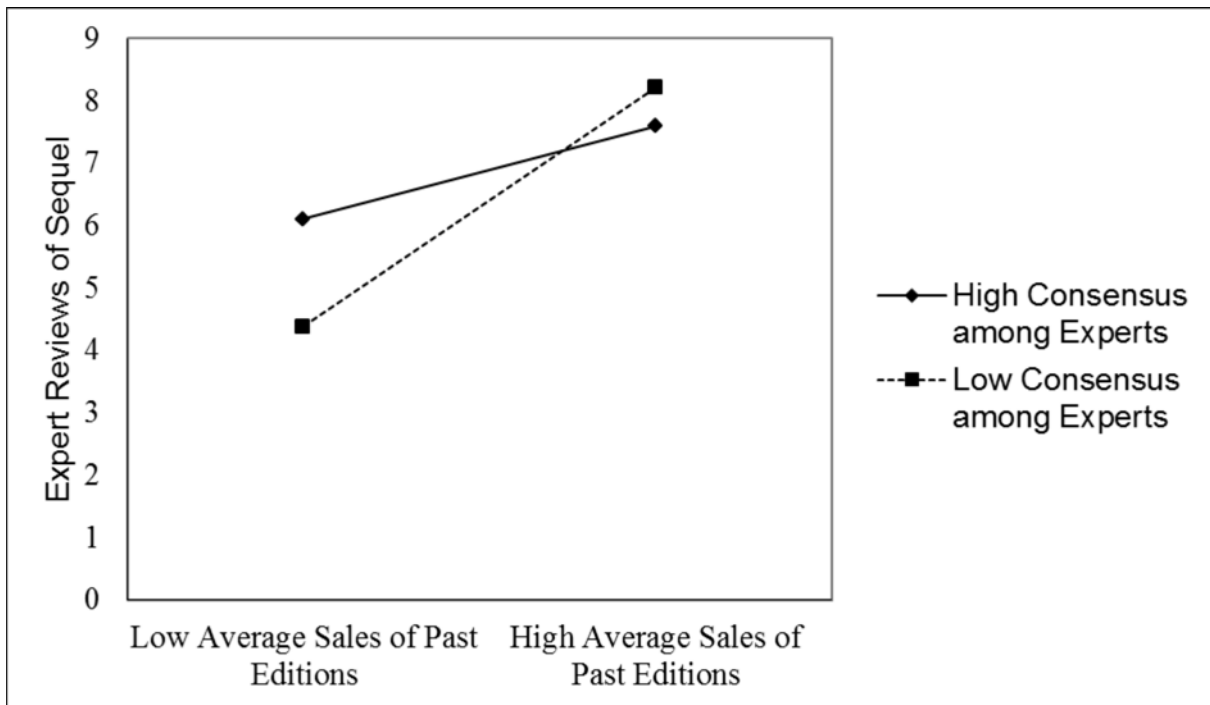


Figure 4. The effect of sales of past editions on expert reviews of the new edition (sequel) with low and high variability of expert reviews