Analysing pre-release consumer buzz and information cascades within the film industry: are there differences by gender and age groups?

Published in the Journal of Media Economics.

Citing this article.

To cite this article: Izquierdo-Sanchez, S & Shaw, A (2022): Analysing pre-release consumer buzz and information cascades within the film industry: are there differences by gender and age groups? *Journal of Media Economics* (In Print).

Corresponding author's info:

Sofia Izquierdo-Sanchez is a senior lecturer in Economics at the University of Huddersfield. She is also a Co-Director of the Northern Productivity Hub. E-mail: S.Izquierdo-Sanchez@hud.ac.uk

Alan Shaw is a senior lecturer at Leeds Beckett University. His key research interests are changing individual and societal behaviours using social marketing. E-mail: Alan.Shaw@leedsbeckett.ac.uk

Abstract

The concept of pre-release consumer buzz (PRCB) is a relatively new phenomenon, it is the excitement generated by consumers in anticipation of a forthcoming new product, film, song, or play. This PRCB is closely associated with information cascades because the buzz generated can be a mechanism for driving consumers to experience the said new product. Earlier research has called for scholars to test the pervasiveness of the concept, there is also concern that current studies only adopt a national overview. We have addressed these concerns by using a large original dataset, collected weekly for approximately one year. We analyse the determinants of information cascades and PRCB by considering films premiered in the USA and the UK. More specifically we examine online user ratings by differing demographic clusters of the population (by sex and age) and through the qualitative characteristics of films (i.e., genre). Our results demonstrate that males between the ages of 18–29-years are more compliant to information cascades and expert reviewers are more likely to instigate herding behaviour.

Introduction

It has been argued that the success of a new product will depend not only on its premarketing but the associated pre-release consumer buzz that is generated (PRCB) (Houston et al., 2018). This PRCB is defined as "the aggregation of observable expressions of anticipation by consumers for a forthcoming new product" (Houston et al., 2018, p.339). The film sector has a particular affinity with this proposition because consumers cannot determine their quality prior to purchase, and increasingly, online review scores for films have been used by potential consumers as a signal of the movie quality (Henning-Thurau et al., 2006). That said, the PRCB concept is relatively new, with Houston et al. (2018) urging scholars to build on their conceptual model to enhance the general theory of buzz. We have accepted this challenge and have embarked on a journey to examine how social learning theory can impact PRCB and thus the general theory of buzz. Social learning theory occurs within a social context and indicates that people learn from one another. Examples of social learning are herd behaviour and information cascades, which occur when agents switch their behaviour to follow the crowd. It should also be noted that although the definitions of information cascades and herd behaviour are very similar and often interchangeable (see Badeley, 2013; Liu et al., 2019; Liu et al., 2021), Smith and Sorensen (2000) emphasize that there is a difference between them. They argue that an information cascade occurs when individuals completely ignore their private beliefs or information to follow the crowd, but herd behaviour occurs when individuals make an identical decision but not necessarily through ignoring their private information. This means that, ultimately, the outcome of acting in herd behaviour or in an information cascade will be the same, with people tending to follow the crowd. We argue that the information cascade is the cause and herding is the effect. Our proposition for this study is that information cascades, which includes PRCB, within the film industry will cause herding. This will be moderated by gender and age differences. As it stands there is only one article directly linking information cascades to PRCB (see, Hennig-Thurau et al. (2012), it focused on how professional reviewer's ratings were associated with box office successes. We aim to extend this and contribute to the general theory of buzz and the information cascades domain by addressing some of the concerns raise Houston et al. (2018) and Hennig-Thurau et al. (2012). The first is where we examine the pervasiveness of PRCB by segmenting the market into gender groups (male/female) and age groups (to test the propositions four were selected: Under 18s, 18-29, 30-44, and 45+). This allowed us to establish how different segments follow the crowd during the buzz phase. It also addresses Hennig-Thurau et al.'s (2012) concerns about how a professional reviewer's ratings are assimilated through different groups.

Our second contribution extends Houston et al. (2018) multi-behaviour approach. We created a large original database, which was built on weekly data for every film released in the UK over a one year[1] period. It included a range of variables, user and expert ratings, director, release dates, cast, advertising expenditure, genre, screens, and budgets. This larger dataset allowed us to control for the quantitative and qualitative characteristics of films. More importantly, we also consider the fact that some films have an extended PRCB because of the rolling launches across different countries, i.e., our dataset also compares films that have their global premiere in the USA before the UK and those that are premiered in the UK first. This also addresses Hennig-Thurau et al.'s (2012) concern that studies should look beyond the USA.

1. Literature Review

The concept of an information cascade was first introduced by Bikhchandani, Hirshleifer, and Welch (1992). They explain that an information cascade occurs when individuals stop believing in their private preferences to follow the behaviour of others. Since this first article,

[1] We collected weekly information for every film released in the UK for one year, 52 weeks. After the year finished we continued collecting information for those films in our database which were still on screens until they were not showing in cinemas, 56 weeks in total.

information cascades have become one of the most popular topics of behavioural economics. Its premise is to understand the economic decisions of individuals so that institutions can adapt their strategies to maximise revenues. Most papers on this topic analyze or develop binary models, here an individual will have two options, i.e., either adopting or rejecting a product and/or service (Bikhchandani et al., 1992, Bikhchandani et al. 1998, Banerjee 1992, Guarino et al. 2011, Welch 1992). Such models assume homogenous preferences and perfect information about the people making decisions ahead of them. More specifically, it is the knowledge of the order in which people have chosen a product or service and of all the past actions of others who have decided before them (Bikhchandani et al. 1998, Banerjee 1992). A few authors tried to analyze information cascades in a context where agents can choose among several options (De Vany and Lee 2000), preferences are heterogeneous (Smith and Sorensen 2000), there is no decision order (Guarino et al. 2011) and only the average of past actions is known (Acemouglu et al. 2011). Ultimately, these authors could not find empirical evidence which supported these assumptions, so they conclude that information cascades were very difficult to identify empirically under such assumptions.

Cascades can explain the rapid process by which society switches from one equilibrium to another, but these movements can be fragile (Bikhchandani, Hirshleifer, and Welch 1992 and 1998, Bowden and McDonald 2008, Russell 2012, Goettle and Phillip 2005) and very often in the wrong direction[2] (Bikhchandani, Hirshleifer, and Welch 1992, Banerjee 1992, Ellis and Fender 2011), due to the fact that cascades start on the basis of a small amount of information. Indeed, Bikhchandani, Hirshleifer, and Welch (1998) and Banerjee (1992) argue that a cascade can arise after three or more individuals have made their selection. What we are now seeing are studies linking and comparing marketing data (i.e., the creation of buzz) with this e-word of mouth (see Houston et al., 2018).

[2] The final outcome does not always maximize individuals' utility.

The existence of information cascades was demonstrated and studied in other fields such as finance (Genesove and Meyer 2001, Welch 1992), firm behaviour (Kennedy 2002) and neuroscience (Gray 2011). Information cascades and/or herd behaviour are likely to occur in several industries but in relation to 'new experience products' we believe the association to be interesting. To explain this, we must define experience products, Nelson (1970) sees it as goods whose quality cannot be determined prior to purchase. Consequently, consumers may use the purchases of others to guide the buying decision; for this reason, the usefulness of advertising and the impact of information cascades will arguably be an important factor in the sales growth of products and services. An example of an experience product is a film. Before consumers go to the cinema, they will not know the quality of the film. It is only after watching the film that the consumer discovers if the film was worth the money. These experience products also include books, video games, music, restaurants, magazines, and plastic surgery.

As highlighted earlier through the work of Houston et al. (2018), the propagation of information that allows information cascades to arise may be driven by external sources such as television, radio, internet commercials or e-word of mouth conversations through social media. In the last few years, the topic of information cascades has attracted the attention of researchers due to the emergence of the internet and online social networks (Acemoglu et al. 2011, Bowden and McDonald 2008, Iribarren and Moro 2011). Research relating to PRCB has been sparse, table 1 provides a summary since the publication of Houston et al. (2018), it also includes a summary of those works within the information cascade and herding domains that call for scholars to embark on studies that include a heterogeneous element their research.

Table 1: A literature review summary of PRCB (since Houston et al. 2018) and Information Cascades/Herding call for more heterogeneous research.

Authors	Definition/ Description of …	Operationalisation	Related Theory or Argument	Data & Methodology	Context /Industry	Related Findings
Raafat et al., 2009	Herding is a form of convergent social behaviour that can be broadly defined as the alignment of the thoughts or behaviours of individuals in a group (herd) through local interaction and without centralized coordination.	Convergence upon a single mood or emotion can elicit herd behaviour in which the agents are connected and process stimuli in a similar manner.	Diffusion Theory/ Social Network Analysis.	Conceptual. Homogeneous.	Happiness /Emotions	It presents scaffolding for organizing the questions that can be addressed about herding.
Baddeley, 2013	Individuals form probabilistic judgments about the likelihood of an event based on others' choices	Deepening the understanding of expert opinion formation and identifying the factors leading to inefficient outcomes in terms of the accumulation of knowledge and learning.	Heuristics, biases and prospect theory.	Herding, Conceptual.	Academic / Expert Opinions.	The externalities from supporting a consensus view are negative, and therefore, there will be over reliance on the opinions of others.
Lee et al., 2015	An individual can rationally follow the behavior of preceding individuals without regard to his/her own information, having observed the actions of those ahead of his/her.	Understanding the social drivers of a user's rating can help managerial practices such as tailored marketing strategy and reliable design of recommender system.	Info Cascades. Online learning, Herding behaviour.	Analysis of USA dataset. Logistic Regression. Homogeneous.	Movies.	Observational learning by others' ratings can trigger herding behavior when a subsequent generates movie rating.
Schauerte et al., 2018	As per Houston et al. (2017)	Comparing linear and digital based TV.	Resource based theory. Buzz.	Conceptual. Homogeneous.	TV.	Value propositions can still keep linear TV organisations relevant.
Liu et al., 2019	An individual, having observed the actions of those ahead of him, to follow the behavior of the preceding individual without regard to his own information.	Content generators require a better understanding of the effects of book ranking and online user reviews on the clicks of e-books.	Information Cascade Theory, E-WOM and Ranking systems.	Information Cascades, 25-week panel dataset on Yuedu.163.com. Bayes Law. Regression Analysis.	Online Reading.	Informational cascades are particularly prominent on the online reading market.

Saunder et al., 2019	In an effort to minimize uncertainty about their own decisions, people tend to utilize information provided by others and converge on similar behaviors.	Actionable guidance for managers concerned with online reputation management, product strategy and planning, and the design of online rating platforms.	Theory of Herding. Herding from reference groups.	Linear Regression of 44,108 individuals with 2,218,574 ratings. Heterogeneous: friends and crowds.	Board Gaming	Experienced raters coalesce more on friends' rating more than with the crowd.
Schaer et al., 2019	The aggregate anticipation of consumers towards a new product	Demonstrating the importance of forecasting the life-cycle sales of new products prior to launch.	Product adoption / augmenting information. Buzz.	Empirical experiment. Homogeneous.	Video games.	Pre-release buzz contains predictive information up to 17 weeks prior to release and can increase life-cycle sales forecast accuracy up to 20%.
Shi et al., 2020	Listening to online consumer activity to predict new product success is becoming increasingly popular.	Demonstrating the importance of buzz in the movie sector.	Purchase intentions. Buzz.	Survey (1500 students) and social media data. Homogeneous	Movies.	Customer buzz based forecasts outperform surveys under conditions of high uncertainly, e.g., for niche and low-budget movies
Liu et al., 2021	An individual, having observed the actions of those ahead of him, to follow the behavior of the preceding individual without regard to his own information.	Online retailers should place a high value on the role of product rankings on consumers' online shopping behavior.	Cultural orientation and Information Cascades.	Information Cascades. Bayes Law. Regression Analysis.	Online Shopping.	Information cascades moderate the impact of price discounts on online purchase behavior, which is also influenced by the cultural orientation of online customers.

These days, it is standard practice for e-commerce sites to provide consumer preference ratings of products. We can also see newspapers and magazines providing expert opinions about products and services. All of this can be considered as a proxy for 'word of mouth' dissemination (Dellarocas et al. 2004) or PRCB (Houston et al., 2018). This feature is very important because literature has shown that the recommendations of other people are very useful for purchases of experience goods such as films or restaurants (Nelson 1970; Zufryden 1996; Lee et al. 2009; Anderson and Madruger 2012). However, information cascades theory not only implies that information is transmitted by the consumers, but it also implies that some individuals, who are called "fashion leaders" (Banerjee 1992) or "royal family" (Bala and Goyal 1998), influence consumer decisions straight away and can trigger immediate cascades. In the case of the film industry, expert critics would be a good example of this (Hilger et al. 2011; Reinstein and Snyder 2005).

The film industry offers a particularly interesting example of an industry in which the phenomenon of information cascades or herd behaviour can be examined (De Vany and Walls 1996; Walls 1998; De Vany and Lee 2000; Moretti 2010). Films can be considered as experience goods since consumers cannot determine their quality prior to purchase, and, increasingly, online review scores for films have started to appear and potential consumers use them as a measure of film quality (Henning-Thurau and Houston 2006). Previous information cascades in the film industry have focused on the autocorrelation of box office revenues, and small datasets that only include top films or blockbusters (De Vany and Walls 1996; De Vany and Lee 2000; Walls 1998). The communication of 'quality' among individuals who have already seen a film, and the information cascades that may arise thereafter, were the main focus of a small number of recent articles (Moretti 2010; Lee et al. 2015). This shows how the cascades and thus PRCB in the film industry can be found within consumers' opinions when considering the qualitative characteristics of movies, using online reviews as a proxy for word of mouth (Dellarocas et al., 2004). The extraction of data from social networks has allowed us to identify a statistically significant convergence on consumers' opinions considering: (1) clusters and a demographic breakdown of the population (by sex and age), (2) the qualitative characteristics of films (by genre). Furthermore, we will show that cascades start to form before the release of a film highlighting the importance of advertising and word of mouth (i.e., PRCB).

2. The theoretical approach

Our study builds on the work of Bikhchandani, Hirshleifer, and Welch (1992, 1998) and Banerjee (1992) by combining it with the markets and crowd's theory of Easley and Kleinberg (2010). It explains the effect of social networks on information cascades and PRCB using the film industry as a case study. More specifically, we examine how cascades arise, the fragility of cascades and what makes a cascade stop.

Assume that in week *t* there are *J* films where *J=1, 2,..., j* and *N* individuals where *N=1, 2,..., n* and *j*<*n*. Films can be rated by users or experts from 0 to 10, 0 indicating a low quality film and 10 a high quality film. The perceived quality of a film is known by the distributors but unknown to individuals prior to visiting the cinema. After watching a film, each individual faces an 'identical choice decision problem' choosing a rating action $a \in \{0, \dots, 10\}$; for simplicity, in this example, looking at the statistical distribution of users' ratings presented in Figure 1. It is assumed that individuals will consider film j to be a good film if the average rating is equal or greater than 7 (r_1), or a bad (or less good) film if the average rating is less than 7 (r_2). An individual pay-off $\mu(a, w)$ will depend on the realization of a state of the world $w \in (H, L)$ H being a high-quality film, and *L,* a low quality film. So, if a film is actually a high-quality film and an individual n gives a rating greater than 7 to the film, then the payoff will be bigger.

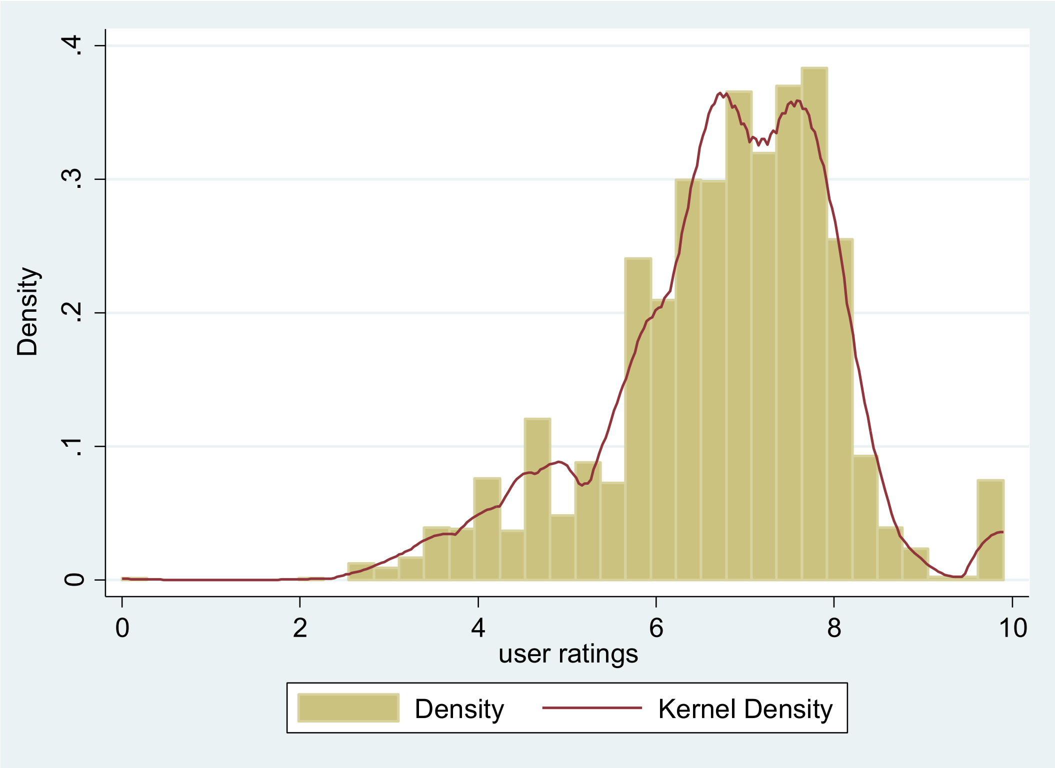


Figure 1: the statistical distribution of users' ratings.

Public information (σ) about a film is available to every individual, but as preferences are heterogeneous each individual considers only a small part of this information (we posit this as their 'private information' or 'private signal'). This private signal can be High (ρ_H) if they consider the film matches their preferences, or Low (ρ_L) if they consider the film doesn't match their preferences. We are assuming that the first individual will make the decision of watching film j based on his/her private signal, however, users' ratings in social networks may start to form before a film is released as some film information such as advertising or trailers are released. Also, some movies are released worldwide while others have different release dates per country, and so potential viewers will already have some rating information available before the release. In terms of the later followers we posit that individuals would only observe the average rating of people ahead of them and the number of people that voted; the rating order is unknown. These ratings affect directly the decisions of potential consumers.

Our dataset has segmented the ratings by age and sex. These ratings are divided into four sub-groups and include the total average rating of the group (under 18s, ages 18-29, ages 30-44 and over 45s), male average ratings (males under 18s, males 18-29, males 30-44 and males over 45s) and female average ratings (females under 18s, females 18-29, females 30-44 and females over 45s). These segments form an interacting network described as follows: each sub-group of ratings such as males under 18, females between 18 and 29, etc. is considered a node, and each group of nodes (total average under 18s, males under 18 and females under 18) will be considered a set of nodes. Based on the work of Easley and Kleinberg (2010) the nodes represent the network structure, i.e., the relationship that each subgroup have with their neighbours. Each individual observes the average rating of each sub-group and can be influenced by them.

Based on the above, this study's conceptual model can be represented by Figure 2. The red discontinuous lines divide each set of nodes into identifying clusters; "a cluster is a set of nodes such that each node in the set has at least a fraction *p* of its network neighbours in the set" (Easley and Kleinberg, 2010 p.500-501), and black continuous lines show peer influence.

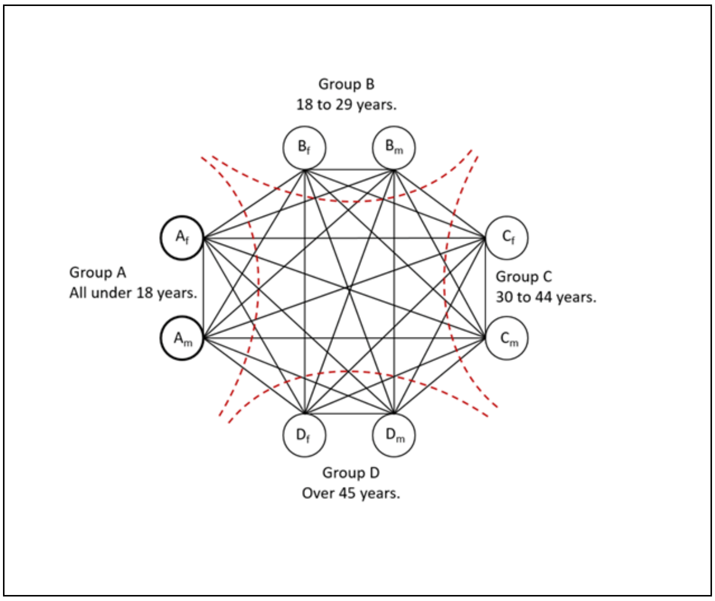


Figure 2: The study's conceptual model.

To articulate this model further, readers must imagine that one cluster has given a high rating (H) to film j, and a low rating (L) to the rest. There is a complete cascade if everyone in the network ignores their private signal to follow consumer's opinion and watch film j. According to the threshold rule, if a fraction of at least $q = \frac{w}{a+w}$ of your neighbours follows one behaviour, then you should too (Easley and Kleinberg, 2010), where the value of q depends on the payoffs a consumer gets from choosing one option or the other, i.e. if participant i gives a rating H the payoff is $a > 0$, and if participant i gives a rating L the payoff is $w > 0$.

In this case, suppose that nodes set A (total average under 18), with node A_m (males under 18), and A_f (females under 18) have adopted option H, and the rest have adopted L; B_m will watch film j if q (threshold point) $\leq 2/8 = 1/4$ (two neighbours have adopted H, with B_m having seven neighbours in total). Assuming that B_m has adopted H, B_f will also decide to watch film j if q $\leq 3/8$ (and since B is the total average rate of B_m and B_f, B will also adopt H). Ratings will continue similarly until they form a complete cascade.

3. Dataset

To address the research questions, a dataset was created for this paper. Weekly data was collected for one year, focusing on every film released in UK cinemas, (a total of 549 films). This dataset included users' and expert critics' scores, director, actors/actresses, advertising expenditure, OSCAR academy award and BAFTA prizes, genre, distributor, number of screens, British Broad of Film Classification (BBFC), box office revenues in the UK and USA, and dummy variables identifying if the film is a sequel or if the original idea comes from a book or comic. User online ratings were collected from The Internet Movie Database (IMDb.com). IMDb is one of the most popular entertainment web pages used across the world

(BBC, 2010)[3], including in the UK. IMDb offers a rating scale from 1 to 10 that allows users to rate films. These scores are presented according to demographic breakdowns, dividing people who provide ratings by gender and age. IMDb provides the cumulative average ratings and the number of voters for each film, i.e., the average of week 3 for film *j* includes the average of week 2 and 1 for the same film, which could lead to problems of heterogeneity in our model. To solve this problem, the individual average rating for each film by week was obtained.

Information regarding film directors, producers, writers, casts, and release dates was collected from Box Office Mojo (boxofficemojo.com). Expert critics' film review scores were collected from different sources, including the following popular British newspapers and websites: The Guardian, Daily Express, Daily Mail, Daily Mirror, Daily Telegraph, The Independent, The Times, The Sun, Yahoo![4], and IMDb Staff critics. The motive for using different expert critics' film review scores was to cover a broad range of potential opinion influencers, who would encourage different groups of people. UK Box Office revenues were provided weekly and extracted from the British Council Film web page. Advertising expenditure was provided by Nielsen, a major international market research company (Elliott and Simmons, 2011).

We include a comprehensive set of variables to control for public information. This information is freely available to anyone, but as preferences are heterogeneous, we use the assumption that individual considerations are only a small part of this data (i.e., the private information or private signal). The public information variables included: directors, release dates, popular actors/actresses, OSCAR and BAFTAs prizes (Elliot and Simmons, 2008; Elsberse, 2007), genre, distributors, budgets, number of screens, British Board of Film Classifications (BBFC), UK box office revenues and expert review scores.

[3] Data was collected by the Nielsen media company in January 2010 measuring webpage usage, covering the UK, France, Germany, Italy, Spain, Switzerland, Brazil, US, and Australia. IMDb is listed number 28 out of the top 100 websites.
[4] Yahoo! is listed number 3 in the BBC (2010) report.

Some of the films had a stock of initial information because they had already been released in the US prior to the UK. We observed that users started rating a film once the trailer was available and/or the film had been released in another country. We studied whether information cascades started forming ahead of the release date and so we also considered movies "USA Premier" and "UK Premier" once they had been released. When studying "USA Premier" movies we controlled for total box office revenues in the US and US expert critics. When studying "UK Premier" movies we control for box office revenues in the UK and UK expert critics (as per Eliashberg and Shugan, 1997). Table 2 provides a description of the variable and Table 3 shows the descriptive statistics information.

Table 2: Description of the Variables

Variable	Description	Definition
Users' ratings	Average ratings on a scale from 1 to 10 divided according to age and gender	females/males under 18 females/males between 18 and 29 females/males between 30 and 44 females/males older than 45
Popular cast	1= Oscar or BAFTA winner nomination in the last 5 years, 0=otherwise	Actor/actress that has been nominated to an Oscar or BAFTA in the last 5 years
Oscar /BAFTA prizes or nominations	1= the film was nominated to, or won an, OSCAR/BAFTA, 0=otherwise	OSCAR/BAFTA nomination or prizes categories 2011/2012
Genre	1 = if a film is classified as comedy, war, action…; 0=otherwise.	
Major Distributor	1 = if the distributor of the film is one of the big 6; 0=otherwise	Big 6 studios: Disney, Century Fox, Sony Pictures, Universal Studios, or Warner Bros
Box Office Revenues UK	Weekly Box Office Revenues in UK (£)	
BBFC under 18	1 = if BBFC (British Board of Film Classification) is General Public (PG), Universal (U), etc. 0= Otherwise	BBFC under 18 = 1 if the BBFC recommends the film is suitable for people under 18 years old; 0 = otherwise.
Advertising expenditure	Weekly advertising expenditure (£)	
UK expert Critics	Weekly UK Expert critics average scores	
Budget	Total production budget	
US Box Office Revenues	Total Box Office Revenues in US ($)	Box office revenues in the US up until the film is released in the UK
US expert critics	US expert critics average scores	

Table 3: Descriptive Statistics

Variable	Mean	Standard Deviation	Min	Max
User ratings				
Females	7.07	1.29	1	10
Males	5.75	0.64	1	10
Aged under 18	7.38	1.58	1	10
Aged 18-29	6.96	0.42	1	10
Aged 30-44	6.51	1.34	1	10
Aged 45+	6.53	1.38	1	10
Votes				
Females	2,882	7,036	1	78,622
Males	11,172	24,805	1	237,763
Aged under 18	704	1,647	1	15,861
Aged 18-29	8,779	19,696	1	194,469
Aged 30-44	3,712	9,066	1	90,976
Aged 45+	893	1,891	1	23,734
Qualitative characteristics of films				
Major distributor	0.46	0.49	0	1
BBFC under 18	0.009	0.097	0	1
Action	0.171	0.376	0	1
Adventure	0.074	0.262	0	1
Animation	0.063	0.243	0	1
Comedy	0.27	0.444	0	1
Crime	0.049	0.217	0	1
Documentary	0.05	0.219	0	1
Drama	0.46	0.209	0	1
Family	0.053	0.224	0	1
Fantasy	0.08	0.282	0	1
Horror	0.05	0.233	0	1
Musical	0.018	0.136	0	1
Romance	0.137	0.344	0	1
Sci-fi	0.06	0.238	0	1
Thriller	0.175	0.38	0	1
War	0.162	0.126	0	1
Popular cast	0.02	0.12	0	1
OSCAR/BAFTA	0.01	0.11	0	1
Quantitative characteristics of films				
Weekly UK box office (£)	2,344,002	242,215.3	20	73,100,000
Budget (£)	18,456,770	3,259,729	0	225,000,000
Weekly advertising expenditure (£)	45,324.9	8,058.6	0	905,996
US box office (£)	20,400,000	8,836,640	0	113,000,000
UK expert critics	6.02	1.75	1	10

Looking at the descriptive statistics in Table 3 we can observe that the average female ratings were slightly higher than the males' average ratings. The age groups that voted more in IMDb were females and males aged between 18 and 29 and between 30-44, against the user aged under 18 and older than 45. The most common genres reviewed were Action, Comedy, Drama, and Thrillers which account for 26%, 30%, 34%, and 18% of the total box office revenues respectively[5]. According to the BFI genre and classification[6] the most common genres of films released in the UK were Comedy and Drama and the highest box office revenues were earned by action movies.

4. Methodology and Results

4.1. Convergence and clusters on user ratings

The dataset provided the cumulative average ratings and the cumulative number of voters for each film; this means that when users provide a rating, they can observe the cumulative average rating per gender and age for film *j* up to week *t*, and similarly users can observe the cumulative number of votes for film *j* up to week *t*. However, this could lead to problems of heterogeneity in the empirical analysis. To solve this problem, the individual average per film for each week and the number of votes per film for each week have been obtained as follows.

$$r_{ijt} = (\hat{r}_{ijt}\ x\ 2) - \hat{r}_{ij(t-1)} \qquad (1)$$

$$v_{ijt} = \hat{v}_{ijt} - \hat{v}_{ij(t-1)} \qquad (2)$$

where r_{ijt} is the average user rating given by group *i* to film *j* in week *t* and $\hat{r}_{ijt}$ is the cumulative average user rating given by group *i* to film *j* in week *t*.

v_{ijt} is the total number of votes by group *i* to film *j* in week *t*, and $\hat{v}_{ijt}$ is the cumulative number of votes by group *i* to film *j* in week *t*.

The question of whether there was a convergence on user ratings was analysed from a statistical perspective, looking at the weighted average and the standard deviation over a given period a movie was shown on screens. Ratings were divided by sex and age as explained in Section 3.

[5] A movie can be classified by the BBF with more than one genre, for example, romantic comedies.
[6] https://www.bfi.org.uk/sites/bfi.org.uk/files/downloads/bfi-genre-and-classification-2017-06-16.pdf

It was noted that, depending on the age or gender, filmgoers could be biased towards specific film genres, teenagers showed a preference for action, and superheroes films (Epstein, 2016). In terms of gender split, females showed a preference for comedies, romance, and drama film genres, while male audiences had a higher preference for action and war movies (BFI, 2016).

Considering the different genres as a way of classifying films with similar public information characteristics, Figures 3 and 4 show the evolution of the weighted average of user ratings over time by sex and age respectively. They show that in all genres, the average ratings of females are higher than males. We were unable to find any literature within the media domain to support or contradict this fact, however there is a large body of work that has focused specifically on gender differences. A notable example is the work of Kajonius and Johnson (2018). They identified that females are more agreeable than males, this may explain why their ratings are higher. The differences in tastes for both genders are again exposed when we look at the weighted standard deviation between females and males' reviews (Appendix 1, Figure A.1), we did however observe a level convergence for some genres such as animation and crime, but in most cases, we can see a divergence of opinion between males and females. Arguably this may be down to females being more predisposed to "tending and befriend rather than [the male] fighting or fleeing" (Eisler & Fry, 2019 p.94). Our study also shows that more males (or more specifically millennial males) are likely to use online rating systems, this supports the findings of Mangold and Smith (2012) and Monaco (2018). It also demonstrates that information cascades are consistent amongst males, but there are greater variations from the female population, particularly in the documentary, animation, and comedy genres.

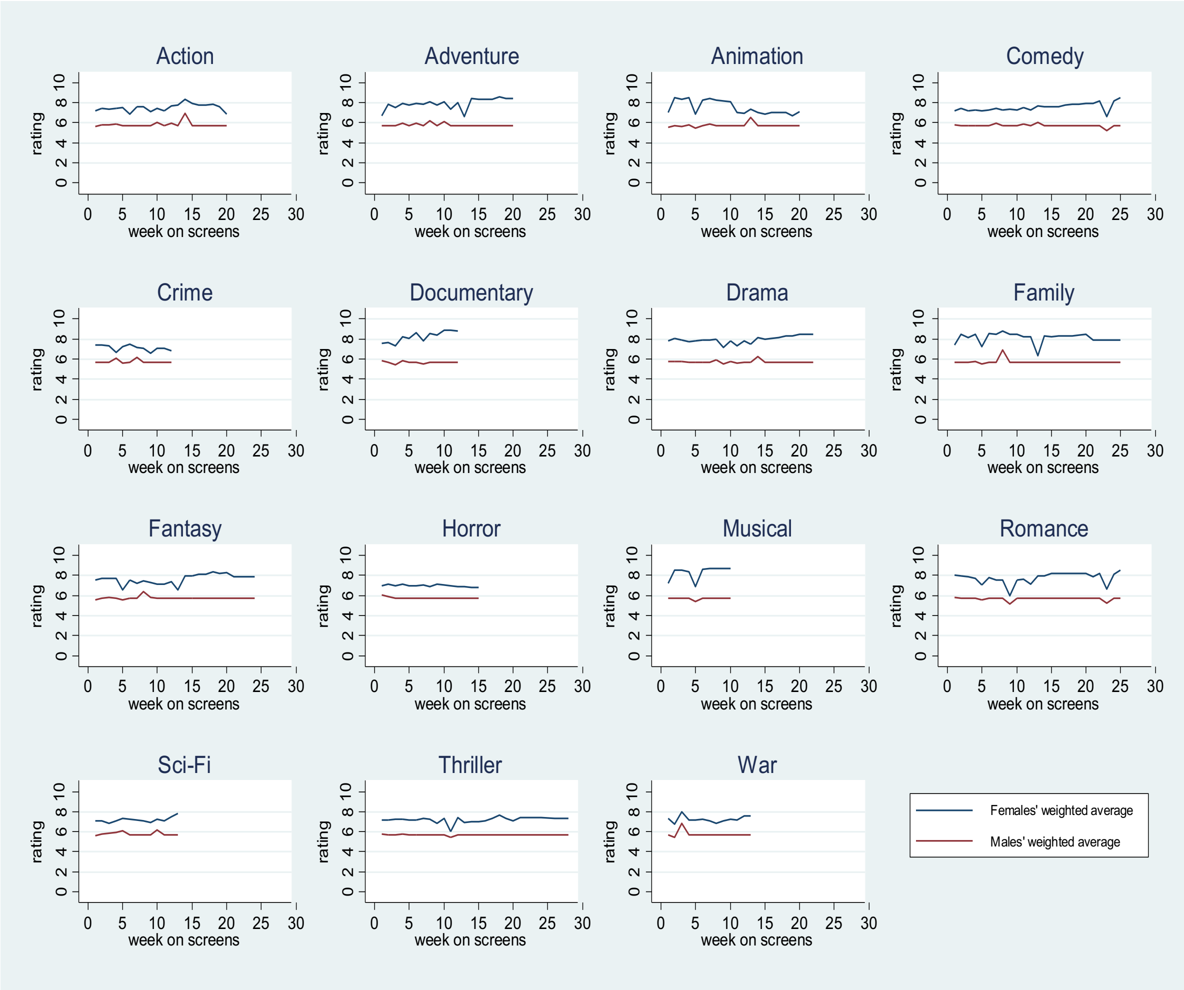


Figure 3: The weighted average of user ratings over time by sex.

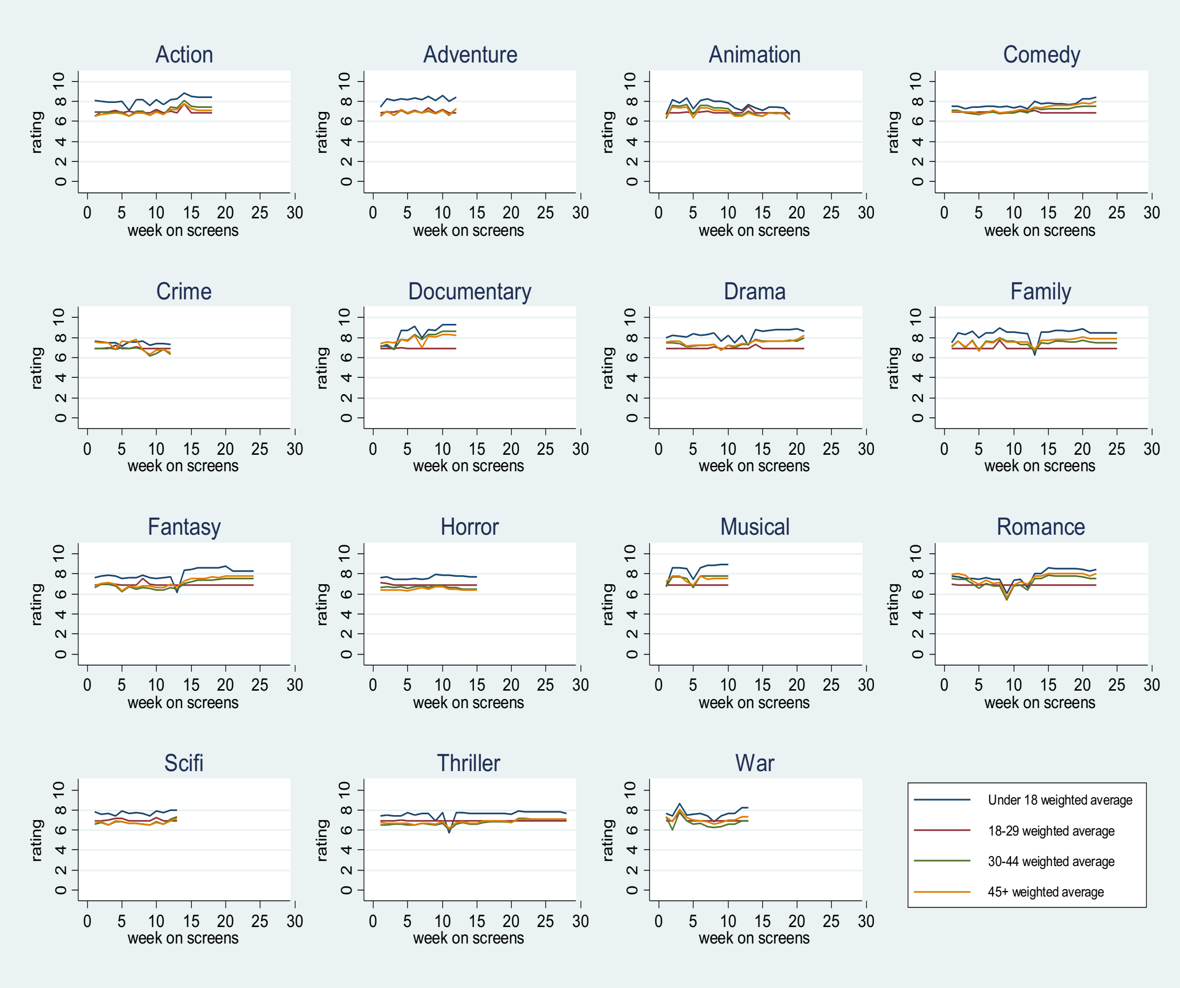


Figure 4: The weighted average of user ratings over time by age.

Analysing the results by age (Figure 4 and A.2 (Appendix 1)), we found that there was a level of convergence between the two younger and the two elder cohorts respectively (see for example: romance, musical) The action genre demonstrated low convergence for all the age cohorts[7]. Although we find divergences across gender and age, we did find similarities within

[7] We confirm these results using a t-test on the equality of means per week for the different genres by age and sex. We map whether rating differences are statistically significant per group and week, using a t-test on the equality of means per week for the different genres. The test is as follows: $H_0: difference_{ijt} = 0$ and $H_A: difference_{ijt} \neq 0$ (1) by gender: $difference_{ijt} = mean\ (females_{jt}) - mean\ (males_{jt})$ (2) by age: $difference_{ijt} = mean\ (group\ aged\ i) - mean\ (group\ aged\ \neq i)$. Results are available to request.

movie genres identifying that different film genres can target similar types of audiences (BFI/Northern Alliance/Ipsos Media C, 2011; Izquierdo Sanchez, 2018; Redfern, 2012). As we alluded to earlier, the variations seen in the age graphs demonstrate that the 18–29-year-olds are more compliant to information cascades. Again, we were unable to find any literature within the media domain to support or contradict this fact, however, considering the work of Monaco (2018), it may be that many of the earlier information cascade studies were inadvertently skewed to participants within the 18–29-year-old segment. Unlike ours, these studies did not control for age, as such we believe that our work adds to the theory and knowledge of the domain by demonstrating who are more likely to use the rating systems. From a practical / managerial point of view this should not be seen as a limitation because building of the work of Caudron and Van Peteghem (2018) these individuals will grow into the other age bands and continue with their digital behaviours.

Before considering the effects of our study we must first explain how the weighted average were calculated:

$$weighted\ average_{ijt} = \sum_{j=1}^{NJ} \alpha_{ijt}\, r_{ijt} \qquad (3)$$

where $\alpha_{ijt} = \frac{number\ of\ votes_{ijt}}{number\ of\ votes_{it}}$ and $\sum_{j=1}^{NJ} \alpha_{ijt} = 1\ for\ all\ t.$

The weighted average is calculated by gender and age.

The standard deviation (SD) is calculated as follows:

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^{NJ} \alpha_{ijt}\, x_{ijt}}{V_{ijt}} \qquad (4)$$

Where $V_{ijt} = \sum_{j=1}^{NJ} \alpha_{ijt} = 1$, and the treatment of x_{ijt}, which represents the rating variance between the groups, will vary as follows:

(1) by gender[8]: $x_{ijt} = (r_{ijt} - average\ rating_{jt})^2$

(2) by age: $x_{ijt} = (r_{ijt} - average\ rating_{jt})^2$

4.2. Information cascades by group-genre in the UK film industry

Following the results in section 5.1, in this instance we divide films in genre cohorts depending on the most likely target audience: Genre group 1: Family, romance, and romantic comedies. Genre group 2: Drama and musical. Genre group 3: Horror, fantasy, comedy, animation, sci-fi, action, and adventure. Genre group 4: Documentary, thriller, and war[9].

We first estimated the following model using an OLS approach:

$$Y_{ijt} = \alpha_0 + \beta_1 \bar{Y}_{ij(t-1)} + \beta_2 \bar{G}_{hj(t-1)} + \beta_3 A_{j(t-1)} + \beta_4 X_{jt} + \beta_5 B_j + \beta_6 U_j + \beta_5 R_{j(t-1)} + \beta_6 K_j + \mu_{ijt} \quad (5)$$

Y_{ijt} is a vector which represents the online rating average for group *i* for film *j* in week *t*. *t=1....56* indicates the time period and *i=1....k* indicates the group studied at time *t*. Being the different groups: Females, Males, Users aged under 18, Users aged 18-29, Users aged 30-44 and Users older than 45 years old.

$\bar{Y}_{ij(t-1)}$ is a vector which represents the average rating of group *i* in week *t-1*.

[8] Note that when calculating the variance per gender there are two groups "females" and "males", this paper uses the weighted standard deviation of females which will give us the convergence with the group males (the convergence result would be the same if the calculation was done with the weighted standard deviation of males).
[9] A film can be classified with more than one genre, in this case film j will be taken into account in all the genre groups (1 to 4) is classified as according to the "target audience" assumption this film will be a potential substitute for other films classified within the same genre groups.

$\bar{G}_{hj(t-1)}$is a vector which represents the average rating of the other groups different to *i* (*h*) in week *t-1*, where groups refer to the different gender and age cohorts.

$A_{j(t-1)}$ is a vector which represents the advertising expenditure observed for film *j* in week *t*, X_{jt} is a vector which represents all the public information available for film *j* in week *t*.

B_j is a vector which represents the total box office revenues earned by film *j* in the US up until the release of film *j* in the UK and U_j is a vector which represents the US expert critics for film *j* in week *t*.

$R_{j(t-1)}$ is a vector which represents one-week lagged box office revenues for film *j* in the UK and K_j which is a vector which represents the UK expert critics for film *j* in week *t*.

Finally, μ_{ijt} is the error term.

The results presented in Table 4 are for the "USA Premiere" movies and Table 5 are for the "UK Premiere" movies, where the dependent variables in columns 1 and 2 are the males and females rating for film j in week t respectively and from column 3 to 6 the dependent variable are the ratings for film j in week t by the different aged groups.

Table 4: OLS USA film premiers

VARIABLES	(1) Males	(2) Females	(3) Under 18	(4) 18-29	(5) 30-44	(6) 45+
Males(t-1)	0.3753***	0.0493				
	(0.098)	(0.069)				
Females(t-1)	0.0909	0.4736***				
	(0.087)	(0.127)				
Under 18(t-1)			0.5965***	0.0634***	0.1273***	0.1035***
			(0.109)	(0.013)	(0.022)	(0.028)
18-29(t-1)			-0.0714	0.5160***	-0.0342	-0.0908
			(0.081)	(0.021)	(0.041)	(0.059)
30-44(t-1)			0.1790**	0.1016***	0.2775***	0.1330*
			(0.089)	(0.022)	(0.064)	(0.077)
45+(t-1)			0.0254	0.0544***	0.1281***	0.4336***
			(0.077)	(0.017)	(0.038)	(0.099)
Popular cast	0.2482	-0.0100	0.2229	-0.0146	-0.0579	0.2450
	(0.208)	(0.131)	(0.365)	(0.088)	(0.154)	(0.155)
BBFC under 18	-0.2403	-0.1275	-0.3890*	-0.2314*	-0.2157	0.1363
	(0.229)	(0.152)	(0.214)	(0.124)	(0.193)	(0.181)
Major distributor	0.0488	0.2274***	0.0364	0.0707*	0.1311**	-0.0499
	(0.084)	(0.071)	(0.112)	(0.039)	(0.064)	(0.085)
log(budget)	0.0597	0.0341	0.1399*	0.0264*	0.0054	0.0196
	(0.039)	(0.039)	(0.072)	(0.015)	(0.033)	(0.057)
log(US Box office revenues(t-1))	-0.0153	0.0055	-0.0405***	-0.0087*	-0.0067	0.0057
	(0.009)	(0.009)	(0.013)	(0.004)	(0.009)	(0.011)
US expert ratings	0.2988***	0.1658***	0.1433***	0.1430***	0.2694***	0.2174***
	(0.037)	(0.032)	(0.046)	(0.015)	(0.033)	(0.040)
log(advertising expenditure(t-1))	-0.0030	0.0146	-0.0042	-0.0042	-0.0053	-0.0061
	(0.011)	(0.009)	(0.018)	(0.006)	(0.013)	(0.013)
Constant	3.4223***	4.4767***	3.0891***	1.4999***	3.0529***	2.7734***
	(0.090)	(0.088)	(0.175)	(0.043)	(0.090)	(0.129)
Genre cohorts	YES	YES	YES	YES	YES	YES
Cluster s.e. (film)	YES	YES	YES	YES	YES	YES
Time dummies	YES	YES	YES	YES	YES	YES
Observations	7,539	7,539	7,522	7,522	7,522	7,522
R-squared	0.675	0.613	0.542	0.899	0.688	0.618

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5. OLS UK film premiers

VARIABLES	(1) Males	(2) Females	(3) Under 18	(4) 18-29	(5) 30-44	(6) 45+
$Males_{(t-1)}$	0.7454***	-0.0465				
	(0.141)	(0.087)				
$Females_{(t-1)}$	-0.6232***	0.2003**				
	(0.139)	(0.083)				
Under $18_{(t-1)}$			0.4546***	-0.0298	-0.1743***	-0.1648***
			(0.102)	(0.024)	(0.058)	(0.060)
$18\text{-}29_{(t-1)}$			0.3435***	0.6484***	0.2497***	0.2252***
			(0.059)	(0.027)	(0.044)	(0.044)
$30\text{-}44_{(t-1)}$			-0.3343**	0.1119**	0.4844***	-0.3075***
			(0.139)	(0.046)	(0.099)	(0.116)
$45+_{(t-1)}$			-0.1597	-0.0592	-0.2398**	0.5661***
			(0.105)	(0.041)	(0.096)	(0.104)
Popular cast	0.1988*	0.1768	0.2175	0.1060*	0.2383**	0.2205**
	(0.115)	(0.125)	(0.132)	(0.059)	(0.105)	(0.106)
BBFC under 18	-0.1694	0.1738	-0.3371*	-0.1504*	-0.0994	-0.0476
	(0.176)	(0.198)	(0.176)	(0.090)	(0.157)	(0.159)
Major distributor	-0.0001	0.3013***	-0.0893	-0.0081	-0.0859	-0.0792
	(0.156)	(0.104)	(0.143)	(0.074)	(0.126)	(0.123)
OSCAR/BAFTA	-0.3837**	0.1701	-0.2043	-0.1886***	-0.0876	-0.0747
	(0.159)	(0.263)	(0.151)	(0.067)	(0.113)	(0.115)
log(budget)	0.0083	-0.0113	0.0215	0.0030	0.0533	0.0871
	(0.076)	(0.040)	(0.065)	(0.033)	(0.058)	(0.058)
log(UK Box office $revenues_{(t-1)}$)	-0.0252	-0.0218	0.0392	-0.0126	-0.0422	-0.0638
	(0.068)	(0.062)	(0.055)	(0.031)	(0.052)	(0.054)
UK expert ratings	0.3723***	0.3042***	0.3085***	0.1696***	0.3460***	0.3392***
	(0.047)	(0.046)	(0.044)	(0.020)	(0.035)	(0.039)
log(advertising $expenditure_{(t-1)}$)	0.0263**	0.0934***	0.0246	0.0124*	0.0087	0.0145
	(0.005)	(0.017)	(0.007)	(0.002)	(0.005)	(0.005)
Constant	4.1989***	6.2463***	2.9060*	1.3486***	2.1608**	2.6220***
	(0.918)	(0.477)	(1.612)	(0.408)	(0.851)	(0.659)
Genre cohorts	YES	YES	YES	YES	YES	YES
Cluster s.e. (film)	YES	YES	YES	YES	YES	YES
Times dummies	YES	YES	YES	YES	YES	YES
Observations	17,933	17,933	17,780	17,819	17,819	17,800
R-squared	0.654	0.453	0.524	0.900	0.692	0.675

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Looking at the results, all lagged group coefficients from the same group were statistically significant and positive, which means that the ratings for each group *i* in week *t* and for film *j* depended on the previous week's ratings for individuals with similar characteristics. Hence, people are influenced by the past actions of their peers not just after the release of a film (see for example: De Vany and Lee, 2000; De Vany and Walls, 1996; Lee et al., 2009; Moretti, 2010; Walls, 1998) but also before the release of a film on screens (Houston et al., 2018).

In both cases (i.e., the "USA Premiere" and "UK Premiere"" movies) we observed that some of the information affecting people's opinions were also statistically significant, suggesting that, before a cascade starts, public information or PCRB is considered, and, as time goes by, consumers tended to follow the crowd but still considered their own private preferences. We did not find that at any point public information was lost "inside the cascade". In this sense, it can be argued that people do make rational decisions by considering other users' opinions, not just making independent decisions by themselves, as would be predicted according to classic economic theories (Tversky and Kahneman, 1986; Simon, 1959). Instead, they decide to follow the lead of other people, even if this decision does not reflect classical utility maximization.

Each group was positively influenced by the opinion of the same aged group but different sex cluster; however, each group also received influences from other groups but the direction of this influence varied by node. Thus, the results highlight the existence and the importance of clusters in the film industry, and the importance of focused advertising by cohorts. Following these results, we can observe social divergence in some groups, as they were negatively affected by other groups due to the high costs of being misidentified with an antagonist group (Bergen and Heath, 2008).

We now calculate the effect sizes of the linear models to look at the strength of the model and the relationship between variables, i.e. "the degree to which the null hypothesis is false" (Cohen, 1977). Following Hays (1963) we use w^2as an index of effect size, this index has previously been used in the context of consumer behaviour (see for example: Peterson et al., 1985 and Steemkamp and Burgess, 2002), and it is equivalent to the r-squared adjusted estimates.

Table 6: Effect sizes for linear model USA Premier. Omega-squared

	(1) **Males**	(2) **Females**	(3) **Under 18**	(4) **18-29**	(5) **30-44**	(6) **+45**
Model	0.676	0.617	0.543	0.899	0.695	0.617
Males(t-1)	0.186	0.003				
Females(t-1)	0.011	0.265				
Under 18(t-1)			0.305	0.049	0.047	0.023
18-29(t-1)			0.003	0.681	0.0015	0.011
30-44(t-1)			0.021	0.071	0.124	0.022
+45(t-1)			0.0004	0.021	0.0311	0.189
Popular cast	0.006	0	0.002	0	0	0.004
BBFC under 18	0.005	0.001	0.006	0.025	0.0044	0.002
Major Distributor	0.0005	0.015	0.00004	0.005	0.005	0.0003
log(Budget)	0.011	0.004	0.025	0.009	0	0.0007
log(US Box office revenues(t-1))	0.013	0.0005	0.338	0.017	0.005	0.0008
US expert reviews	0.372	0.173	0.529	0.374	0.349	0.187
log(advertising expenditure(t-1))	0	0.001	0	0.0003	0.00007	0.00002

Table 7: Effect sizes for linear model UK Premier. Omega-squared

	(1) **Males**	(2) **Females**	(3) **Under 18**	(4) **18-29**	(5) **30-44**	(6) **+45**
Model	0.653	0.452	0.522	0.899	0.691	0.674
Males(t-1)	0.207	0.0003				
Females(t-1)	0.165	0.021				
Under 18(t-1)			0.115	0.003	0.032	0.026
18-29(t-1)			0.15	0.781	0.144	0.109
30-44(t-1)			0.021	0.013	0.075	0.028
+45(t-1)			0.006	0.004	0.025	0.114
Popular cast	0.011	0.004	0.009	0.123	0.02	0.015
BBFC under 18	0.004	0.001	0.014	0.014	0.002	0.0003
Major Distributor	0	0.002	0.002	0.00002	0.003	0.002
OSCAR/BAFTA	0.013	0.009	0.003	0.013	0.0009	0.0005
log(Budget)	0.00009	0	0.0004	0.00006	0.006	0.013
log(UK Box office revenues(t-1))	0.002	0.0008	0.003	0.002	0.226	0.011
UK expert reviews	0.312	0.237	0.18	0.269	0.329	0.296
log(advertising expenditure(t-1))	0.007	0.004	0.004	0.007	0.001	0.002

The columns in Table 6 and 7 refer to the linear models in Table 4 and 5 respectively. If we look at the results in table 5 column 1, we observe that the overall omega-squared indicates that our model accounts for approximately 68% of the variability and the lagged review from the same group (Males) accounts for around 19%. The largest effect comes from the US expert reviews explaining almost 37% of the variability and highlighting the idea of influencing user reviews at an early stage by experts (Bala and Goyal 1998; Banerjee 1992). In fact, we look at the effect sizes of every model we can see that the largest effect always comes from the expert reviews, US expert reviews before the movies have been released in the UK and UK expert reviews after the films have been released in the UK: demonstrating the importance of PCRB, that said there will only be a positive information cascade if the expert reviews are positive. This result highlights the power that the press and "experts" have on consumer behaviour (Hilger et al. 2011; Reinstein and Snyder 2005) and although this effect is combined with the influence from the society and, in the case of the film industry, other public information available (as shown in the results above), the effect size of the expert reviews is still prevalent and bigger than the others. Table 6 and 7 also support our earlier findings that the 18-29-year-olds are more likely to comply with information cascades: they explain 68% and 78% variability respectively in the "USA Premiere" and "UK Premiere" information cascades.

4.3. Robustness Checks

To analyze the robustness of the results we estimate equation (1) using film fixed effects:

$$Y_{ijt} = \alpha_j + \beta_1 \bar{Y}_{ij(t-1)} + \beta_2 \bar{G}_{hj(t-1)} + \beta_3 A_{j(t-1)} + \beta_4 X_{jt} + \beta_5 B_j + \beta_6 U_j + \beta_5 R_{j(t-1)} + \beta_6 K_j + \mu_{ijt} \quad (6)$$

Where α_j is are the film fixed effects. The movie fixed effects are related to the qualitative characteristics of films invariant with time.

Table 8. Film fixed effects USA film premiers

VARIABLES	(1) Males	(2) Females	(3) Under 18	(4) 18-29	(5) 30-44	(6) 45+
Males$_{(t-1)}$	0.0043***	0.0170***				
	(0.001)	(0.003)				
Females$_{(t-1)}$	0.0069***	0.0169***				
	(0.001)	(0.002)				
Under 18$_{(t-1)}$			0.0086*	-0.0022***	-0.0102***	0.0288***
			(0.005)	(0.001)	(0.001)	(0.002)
18-29$_{(t-1)}$			-0.0113**	0.4992***	-0.0025*	-0.0264***
			(0.005)	(0.001)	(0.002)	(0.003)
30-44$_{(t-1)}$			0.0297***	0.0064***	0.0050***	0.0163***
			(0.006)	(0.001)	(0.002)	(0.003)
45+$_{(t-1)}$			0.0190***	0.0097***	0.0038**	0.0350***
			(0.006)	(0.001)	(0.002)	(0.003)
log(US Box office revenues(t-1))	0.0012	-0.0032		-0.0029		0.0051
	(0.013)	(0.023)		(0.011)		(0.036)
log(advertising expenditure$_{(t-1)}$)	0.0011	0.0021	-0.0058**	0.0008	0.0025**	-0.0001
	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.002)
Constant	6.6874***	7.2250***	7.3928***	3.5194***	6.4620***	6.6165***
	(0.005)	(0.011)	(0.022)	(0.007)	(0.011)	(0.022)
Time fixed effects	YES	YES	YES	YES	YES	YES
Number of films	469	469	406	406	406	406

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 9. Film fixed effects UK film premiers

VARIABLES	(1) Males	(2) Females	(3) Under 18	(4) 18-29	(5) 30-44	(6) 45+
Males$_{(t-1)}$	0.0328***	-0.1463***				
	(0.002)	(0.003)				
Females$_{(t-1)}$	-0.0356***	0.1261***				
	(0.002)	(0.003)				
Under 18$_{(t-1)}$			0.1164***	0.0085***	0.0407***	-0.1096***
			(0.004)	(0.001)	(0.002)	(0.003)
18-29$_{(t-1)}$			-0.0042	0.4978***	0.0060***	-0.0089***
			(0.003)	(0.001)	(0.001)	(0.002)
30-44$_{(t-1)}$			-0.0711***	-0.0178***	0.0290***	-0.0733***
			(0.006)	(0.002)	(0.003)	(0.005)
45+$_{(t-1)}$			-0.0738***	0.0062***	-0.0838***	0.2020***
			(0.006)	(0.002)	(0.003)	(0.004)
log(UK Box office revenues$_{(t-1)}$)	0.0206***	0.0950***	0.0379***	-0.0166***	0.0919***	0.0828***
	(0.004)	(0.006)	(0.010)	(0.004)	(0.005)	(0.008)
log(advertising expenditure$_{(t-1)}$)	0.0006	0.0039***	-0.0126***	0.0017***	0.0048***	-0.0015
	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
Constant	7.1968***	6.9156***	8.4128***	4.0579***	6.1910***	5.0011***
	(0.050)	(0.080)	(0.135)	(0.048)	(0.065)	(0.098)
Time fixed effects	YES	YES	YES	YES	YES	YES
Observations	13,451	13,364	13,086	13,517	1,263	12,998
R-squared	0.198	0.193	0.131	0.913	0.205	0.211
Number of titlen	391	391	354	381	380	374

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In general, differences can be found in the public information or preferences considered, emphasizing the idea of the difference in consumer behaviour and tastes between males and females, and between the different age groups. This information can be very useful for producers and distributors when selling or advertising a film, as they could classify the film, identify their objective group, and emphasize, using advertising, the strong points that affect this group, increasing the positive opinions and word of mouth about a film, and thus affecting cinema attendance.

Following equation (5) we now consider the interaction variables of gender and age. Specifically, we estimate the following equation:

$$Gender * Age_{ijt} = \alpha_0 + \beta_1 \overline{Gender * Age}_{ij(t-1)} + \beta_2 \overline{Gender * Age}_{hj(t-1)} + \beta_3 A_{j(t-1)} + \beta_4 X_{jt} + \beta_5\, B_j + \beta_6 U_j + \beta_5\, R_{j(t-1)} + \beta_6 K_j + \mu_{ijt} \qquad (7)$$

$Gender * Age_{ijt}$ being the different groups: Females under 18, Females aged 18-29, Females aged 30-44, Females older than 45 years old, Males under 18, Males aged 18-29, Males aged 30-44, and Males older than 45 years old

Table 10a: Female-Age interactions USA film premiers

VARIABLES	(1) Females under 18	(2) Females 18-29	(3) Females 30-44	(4) Females 45+
Females under $18_{(t-1)}$	0.5257***	0.0158	-0.0052	0.0052
	(0.116)	(0.035)	(0.038)	(0.025)
Females $18\text{-}29_{(t-1)}$	0.2025*	0.3661***	0.0624	0.1110
	(0.110)	(0.114)	(0.052)	(0.074)
Females $30\text{-}44_{(t-1)}$	0.0448	0.0335	0.4269***	0.0748*
	(0.070)	(0.044)	(0.096)	(0.040)
Females $45+_{(t-1)}$	-0.0454	0.0653**	0.0792*	0.7633***
	(0.066)	(0.032)	(0.043)	(0.082)
Males under $18_{(t-1)}$	0.0901	0.0686***	0.0408	0.0749*
	(0.060)	(0.025)	(0.026)	(0.040)
Males $18\text{-}29_{(t-1)}$	-0.0422	-0.0207	0.0044	-0.0105
	(0.123)	(0.060)	(0.062)	(0.126)
Males $30\text{-}44_{(t-1)}$	-0.1987	0.0426	-0.0243	-0.1553
	(0.149)	(0.055)	(0.087)	(0.127)
Males $45+_{(t-1)}$	0.1067*	-0.0483	-0.0135	-0.0545
	(0.059)	(0.034)	(0.044)	(0.107)
Popular cast	0.2522	-0.0027	-0.1997	-0.0486
	(0.317)	(0.170)	(0.216)	(0.164)
BBFC under 18	-0.5059*	-0.1747	0.3239*	-0.3545
	(0.269)	(0.133)	(0.194)	(0.242)
Major distributor	0.1145	0.2398***	0.1212	-0.0062
	(0.106)	(0.077)	(0.080)	(0.081)
log(budget)	0.0077	0.0602	-0.0047	0.0197
	(0.064)	(0.037)	(0.046)	(0.059)
log(US Box office revenues(t-1))	-0.0095	-0.0134	0.0400***	0.0053
	(0.015)	(0.008)	(0.013)	(0.014)
US expert ratings	0.1459**	0.2244***	0.1443***	0.1433***
	(0.057)	(0.030)	(0.039)	(0.049)
log(advertising expenditure(t-1))	0.0011	0.0007	-0.0091	0.0020
	(0.017)	(0.009)	(0.019)	(0.012)
Genre cohorts	YES	YES	YES	YES
Cluster s.e. (film)	YES	YES	YES	YES
Time dummies	YES	YES	YES	YES
Constant	6.6972***	5.5536***	5.5089***	5.5858***
	(0.707)	(0.422)	(0.474)	(0.713)
Observations	7,373	7,373	7,373	7,373
R-squared	0.426	0.639	0.540	0.584

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 10b: Male-Age interactions USA film premiers

VARIABLES	(1) Males under 18	(2) Males 18-29	(3) Males 30-44	(4) Males 45+
Males under $18_{(t-1)}$	0.5506***	0.0817**	0.0810***	0.0286
	(0.075)	(0.033)	(0.025)	(0.023)
Males $18\text{-}29_{(t-1)}$	0.0791	0.2318**	0.0300	-0.1222
	(0.081)	(0.098)	(0.055)	(0.096)
Males $30\text{-}44_{(t-1)}$	0.0742	0.2074***	0.3547***	0.2295***
	(0.073)	(0.051)	(0.057)	(0.042)
Males $45+_{(t-1)}$	-0.0096	-0.0068	0.0540	0.3553***
	(0.042)	(0.032)	(0.035)	(0.079)
Females under $18_{(t-1)}$	-0.0329	-0.0285	-0.0322	0.0097
	(0.046)	(0.034)	(0.034)	(0.028)
Females $18\text{-}29_{(t-1)}$	0.0251	0.0410	0.0084	0.0304
	(0.071)	(0.037)	(0.041)	(0.070)
Females $30\text{-}44_{(t-1)}$	-0.1213**	-0.0135	-0.0202	-0.0333
	(0.048)	(0.030)	(0.024)	(0.045)
Females $45+_{(t-1)}$	0.1247***	-0.0215	-0.0007	0.0529
	(0.044)	(0.023)	(0.028)	(0.033)
Popular cast	0.5472**	0.0605	0.0864	0.2865
	(0.269)	(0.208)	(0.179)	(0.243)
BBFC under 18	-0.3223	-0.4192*	-0.2639	0.2143
	(0.232)	(0.240)	(0.196)	(0.193)
Major distributor	0.0764	0.0796	0.0745	-0.0540
	(0.122)	(0.081)	(0.074)	(0.093)
log(budget)	0.1756**	0.0679**	0.0159	0.0317
	(0.084)	(0.032)	(0.040)	(0.062)
log(US Box office revenues(t-1))	-0.0488***	-0.0093	0.0016	0.0090
	(0.015)	(0.008)	(0.008)	(0.012)
US expert ratings	0.2055***	0.3096***	0.3310***	0.2525***
	(0.052)	(0.031)	(0.034)	(0.042)
log(advertising expenditure(t-1))	-0.0028	-0.0087	-0.0005	-0.0029
	(0.015)	(0.013)	(0.010)	(0.013)
Genre cohorts	YES	YES	YES	YES
Cluster s.e. (film)	YES	YES	YES	YES
Time dummies	YES	YES	YES	YES
Constant	5.1392***	4.3098***	4.0062***	4.0113***
	(0.923)	(0.391)	(0.496)	(0.730)
Observations	7,373	7,373	7,373	7,373
R-squared	0.568	0.724	0.732	0.623

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 11a: Female-Age interaction UK film premiers

VARIABLES	(1) Females under 18	(2) Females 18-29	(3) Females 30-44	(4) Females 45+
Females under $18_{(t-1)}$	0.6677***	-0.0896*	-0.0627	-0.0812
	(0.064)	(0.047)	(0.057)	(0.056)
Females $18\text{-}29_{(t-1)}$	-0.5691***	0.3715***	-0.5315***	-0.5366***
	(0.195)	(0.141)	(0.182)	(0.195)
Females $30\text{-}44_{(t-1)}$	-0.0480	-0.1335	0.7382***	-0.1289
	(0.136)	(0.113)	(0.125)	(0.125)
Females $45+_{(t-1)}$	-0.0232	-0.0807	-0.0532	0.7942***
	(0.099)	(0.055)	(0.087)	(0.087)
Males under $18_{(t-1)}$	-0.0637	0.0076	0.0411	0.0230
	(0.081)	(0.045)	(0.056)	(0.060)
Males $18\text{-}29_{(t-1)}$	0.4897**	0.0398	-0.1199	-0.0611
	(0.205)	(0.121)	(0.176)	(0.189)
Males $30\text{-}44_{(t-1)}$	-0.3352*	0.0619	0.1516	0.1512
	(0.199)	(0.151)	(0.167)	(0.165)
Males $45+_{(t-1)}$	-0.1170	-0.1100	-0.0567	-0.0449
	(0.127)	(0.100)	(0.106)	(0.102)
Popular cast	0.1356	0.0765	0.1223	0.1106
	(0.123)	(0.100)	(0.109)	(0.112)
BBFC under 18	-0.1552	-0.0474	-0.0187	-0.0461
	(0.170)	(0.149)	(0.166)	(0.181)
Major distributor	-0.0925	0.2257***	0.0316	0.0274
	(0.139)	(0.079)	(0.132)	(0.134)
OSCAR/BAFTA	-0.2262	-0.1913	-0.2959**	-0.3615**
	(0.170)	(0.154)	(0.132)	(0.148)
log(budget)	-0.0058	0.0132	0.0082	-0.0004
	(0.065)	(0.036)	(0.056)	(0.063)
log(UK Box office revenues(t-1))	0.0314	-0.0102	-0.0176	0.0044
	(0.056)	(0.054)	(0.050)	(0.055)
UK expert ratings	0.2529***	0.2880***	0.3196***	0.2886***
	(0.043)	(0.039)	(0.044)	(0.044)
log(advertising expenditure(t-1))	0.0283**	0.0559***	0.0063	0.0237*
	(0.013)	(0.012)	(0.012)	(0.014)
Genre cohorts	YES	YES	YES	YES
Cluster s.e. (film)	YES	YES	YES	YES
Time dummies	YES	YES	YES	YES
Constant	5.9766***	5.2346***	5.3839***	4.9364***
	(0.725)	(0.773)	(0.729)	(0.945)
Observations	17,513	38,293	17,628	17,590
R-squared	0.422	0.503	0.609	0.569

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 11b: Male-Age interaction UK film premiers

VARIABLES	(1) Males under 18	(2) Males 18-29	(3) Males 30-44	(4) Males 45+
Males under 18$_{(t-1)}$	0.7681***	0.0384	-0.0205	-0.0015
	(0.102)	(0.063)	(0.061)	(0.058)
Males 18-29$_{(t-1)}$	0.2996	0.7696***	0.2968*	0.2748*
	(0.194)	(0.173)	(0.160)	(0.156)
Males 30-44$_{(t-1)}$	-0.1517	0.1623	0.6140***	-0.1288
	(0.219)	(0.168)	(0.154)	(0.154)
Males 45+$_{(t-1)}$	-0.0861	-0.0537	-0.1010	0.6054***
	(0.120)	(0.104)	(0.103)	(0.106)
Females under 18$_{(t-1)}$	-0.1666***	-0.0902*	-0.1205**	-0.1427**
	(0.064)	(0.054)	(0.052)	(0.055)
Females 18-29$_{(t-1)}$	-0.4919**	-0.4633***	-0.4305**	-0.3641**
	(0.215)	(0.172)	(0.173)	(0.179)
Females 30-44$_{(t-1)}$	-0.1213	-0.1708	-0.0627	-0.0427
	(0.135)	(0.116)	(0.113)	(0.120)
Females 45+$_{(t-1)}$	-0.0302	-0.0888	-0.0403	-0.0447
	(0.092)	(0.078)	(0.071)	(0.079)
Popular cast	0.1634	0.2007*	0.1972*	0.2054**
	(0.126)	(0.117)	(0.102)	(0.103)
BBFC under 18	-0.2406	-0.1602	-0.0629	-0.0016
	(0.196)	(0.184)	(0.156)	(0.159)
Major distributor	-0.0984	0.0001	-0.0711	-0.0612
	(0.161)	(0.145)	(0.128)	(0.125)
OSCAR/BAFTA	-0.4231**	-0.4278***	-0.2866**	-0.1879
	(0.178)	(0.147)	(0.126)	(0.131)
log(budget)	-0.0018	-0.0121	0.0052	0.0567
	(0.071)	(0.063)	(0.059)	(0.061)
log(UK Box office revenues(t-1))	0.0735	0.0121	-0.0138	-0.0733
	(0.065)	(0.059)	(0.056)	(0.057)
UK expert ratings	0.3478***	0.3483***	0.3445***	0.3518***
	(0.047)	(0.047)	(0.038)	(0.039)
log(advertising expenditure(t-1))	0.0310**	0.0281**	0.0193*	0.0174
	(0.015)	(0.013)	(0.011)	(0.012)
Genre cohorts	YES	YES	YES	YES
Cluster s.e. (film)	YES	YES	YES	YES
Time dummies	YES	YES	YES	YES
Constant	2.5745**	4.3398***	3.5712***	4.2153***
	(1.046)	(1.024)	(0.966)	(0.682)
Observations	17,570	17,666	17,647	17,628
R-squared	0.557	0.658	0.687	0.692

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The results emphasize our line of argument in Section 5 users are influenced by their own aged/gender group, however, we can see that their behaviour is also influenced by other groups in either a positive or negative way (social divergence) observing clusters in the film industry. This influence on user's opinions by the society is combined with other information available to consumers and more importantly, we can see the strong positive and significant effect form the expert reviews, how they can consistently affect user's opinions of different gender and ages highlighting the power of the press. Advertising efforts should be then focused on the period before the release of a film but are still effective after the release of the film. However, after the release of a film advertising will have the same impact as positive information cascades.

5. Conclusions

This paper examines consumer behaviour in experience goods, it used the UK film industry as the test case, comparing the ratings of USA Premiere films with UK Premiere films. Information cascades are identified in both cases. We believe that the investigation of social networks allowed us to identify information cascades in the film industry considering (1) clusters and a demographic breakdown of the population (by sex and age), (2) the qualitative characteristics of films (by genre). We identified that females gave, on average, a higher ratings than males but males within the 18-29-year-old age group were more prolific on movie ratings platforms. More importantly it was the 'Expert' reviews that had the biggest effect on the information cascades.

These results contribute to the information cascades literature. Firstly, this paper analysed user ratings before the release of a product. Secondly, an attempt has been made to contribute to the empirical evidence on information cascades literature, which is still limited and characterized so far by a general inability to find empirical evidence of information cascades.

Furthermore, Distributors and producers can use this information to develop a better advertising strategy to sell a film to the general public; efforts should be concentrated on positively affecting the group of interest from an early stage, i.e. before the film opens in theatres, and encouraging word of mouth and cinema attendance.

Declarations of Interest Statement

The authors declare no conflict of interest.

References

Acemoglu, Daron, Munther A. Dahleh, Ilan Lobel and Asuman Ozdaglar. 2011, "Bayesian Learning in Social Networks," *Review of Economic Studies,* 78 (4), 1201-1236.

Andersson, David, E., and Åke E. Andersson. 2006, "*The economics of experiences the arts and entertainment,"* Cheltenham, UK: Edward Elgar Publishing.

Anderson, Michael and Jeremy Magruder. 2012, "Learning from the Crowd: Regression Discontinuity Estimates of the Effects of an Online Review Database," *Economic Journal,* 122 (563), 957-989.

Andersen, Esben S., and Kristian Philipsen 1998, "The evolution of credence goods in customer markets: exchanging "pigs in pokes," In *DRUID Winter Seminar, Middelfart* (Vol. 10).

Arellano, Manuel and Stephen Bond. 1991, "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *Review of economic studies,* 58 (2), 277-297.

Baddeley, M. (2013). Herding, social influence and expert opinion. Journal of Economic Methodology, 20(1), 35-44.

Banerjee Abhijit V. 1992, "A Simple Model of Herd Behaviour," *Quarterly Journal of Economics,* 107 (3), 797-817.

Benz, Men-Andri. 2007, "*Strategies in markets for experience and credence goods,"* Wiesbaden, Germany: Frankfurt Springer Science and Business Media.

Berger Jonah and Chip Heat. 2008, "Who Drives Divergence? Identity Signaling Outgroup Dissimilarity and the Abandonment of Cultural Tastes," *Journal of Personality and Social Psychology,* 95 (3), 593-607.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch. 1992, "A theory of Fads Fashion Custom and Cultural Change as Informational Cascades," *Journal of Political Economy,* 100 (5), 992-1026.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch. 1998, "Learning from the Behaviour of Others: Conformity Fads and Information Cascades," *Journal of Economic Perspectives,* 12 (3), 151-170.

Boes, Stefan and Rainer Winkelmann. 2005, "Ordered Response Models *Working Paper nº 0507,"* Socioeconomic Institute University of Zurich. Available at: http://www.econ.uzh.ch/static/wp_soi/wp0507.pdf (accessed 20 December 2018).

Caudron, J., & Van Peteghem, D. (2018). *Digital transformation: A model to master digital disruption 3rd Ed*. Pennsauken, NJ. BookBaby.

Chevalier, Judith A., and Dina Mayzlin. 2006, "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research,* 43 (3), 345-354.

Clement, Michel, Sibille Fabel and Christina Schmidt-Stolting. 2006, "Diffusion of hedonic goods: A literature review," *The International Journal on Media Management,* 8 (4), 155-163.

Cohen, J. (1977). "Statistical Power Analysis for the Behavioral Sciences", *New York: Academic Press*

De Vany, Arthur and Cassey Lee. 2000, "Quality Signals in Information Cascades and the Dynamics of the Distribution of Motion Picture Box Office Revenues," *Journal of Economic Dynamics and Control,* 25 (3-4), 593-614.

De Vany, Arthur and W. David Walls. 1996, "Bose-Einstein Dynamics and Adoptive Contracting in the Motion Picture Industry," *Economic Journal,* 106 (439), 1493-1514.

Dellarocas, Chrysanthos N., Neveen Awad and Xiaoquan M. Zhang. 2004, "Using Online Reviews as a Proxy of Word of Mouth for Motion Picture Revenue Forecasting." Available at: http://ssrncom/abstract=620821 (accessed 20 December 2018).

Dodds, Klaus. 2006, "Popular geopolitics and audience dispositions: James Bond and the internet movie database (IMDb)," *Transactions of the Institute of British Geographers,* 31 (2),116-130.

Dolgin, Alexander. 2008, *"The economics of symbolic exchange,"* London, UK: Springer Science and Business Media.

Easley, David and Jon Kleinberg. 2010, "*Networks Crowds and Markets: Reasoning about a Highly Connected World,"* New York: Cambridge University Press.

Guarino, Antonio, Heike Harmgartb and Steffen Hucka. 2011, "Aggregate Information Cascades," *Games and Economic Behaviour,* 73 (1), 167-185.

Hays, W. L. (1963). "Statistics for Psychologists", *New York: Holt, Rinehart, and Winston*.

Hennig-Thurau, Thorsten, Mark B. Houston and Shrihari Sridhar. 2006, "Can Good Marketing Carry a Bad Product? Evidence from the Motion Picture Industry," *Marketing Letters,* 17 (3), 205-219.

Hill, Shawndra, Foster Provost and Chris Volinsky. 2006, "Network-Based Marketing: Identifying Likely Adopters via Consumer Networks," *Statistical Science,* 21 (2), 256-276.

Houston, M. B., Kupfer, A-K., Hennig-Thurau, T., and Spann, S. (2018) "Pre-release consumer buzz." *Journal of the Academy of Marketing Science*, 46(2): 338-360.

Izquierdo Sanchez, S. 2019. "Managing the supply of short-life products. A duration analysis approach using the UK film industry", *Bulletin of Economic Research,* 71(1): 75-89.

Izquierdo-Sanchez, Sofia, Caroline Elliott and Robert Simmons. 2016, "The substitution between leisure activities: a quasi-natural experiment using sports viewing and cinema attendance," *Applied Economics,* 48 (40), 3848-3860.

Iyengar, Raghuram, Christophe Van den Bulte and Thomas W. Valente. 2011, "Opinion Leadership and Social Contagion in New Product Diffusion," *Marketing Science,* 30 (2), 195-212.

Kawagoe, Toshiji and Shinichi Sasaki. 2006, "Is ignoring public information best policy? reinforcement learning in information cascades," In C Bruun (Ed) *Advances in artificial economics: The economy as a complex dynamic system.* (pp 191-202) New York: Springer.

Klein, Lisa R. 1998, "Evaluating the potential of interactive media through a new lens: Search versus experience goods," *Journal of Business Research,* 41 (3), 195-203.

Lascu, Dana-Nicoleta. 1991, "Consumer guilt: examining the potential of a new marketing construct," *Advances in Consumer Research,* 18, 290-295

Lee, Young J., Yong Tan and Kartik Hosanagar. 2015, "Do I follow my friends or the crowd? Information cascades in online movie ratings," *Management Science,* 61 (9), 2241-2258.

Liu, Q., Zhang, B., Wang, L., Zhang, X., & Li, Y. (2021). Information Cascades and Online Shopping: A Cross-Cultural Comparative Study in China and the United States. *Journal of Global Information Management (JGIM),* 29(3), 26-45.

Liu, Q., Zhang, X., Zhang, L., & Zhao, Y. (2019). The interaction effects of information cascades, word of mouth and recommendation systems on online reading behavior: An empirical investigation. *Electronic Commerce Research*, 19(3), 521-547.

Moretti, Enrico. 2010, “Social Learning and Peer Effects in Consumption: Evidence from Movie Sales,” *The Review of Economics Studies,* 78 (1), 356-393.

Nakayama, Makoto, Norma Sutcliffe and Yun Wan. 2010, “Has the web transformed experience goods into search goods?” *Electronic Markets,* 20 (3-4), 251-262.

Nelson, Phillip. 1970, “Information and Consumer Behavior,” *Journal of Political Economy,* 78 (2), 311-329.

O'curry Suzanne and Michal Strahilevitz. 2001, “Probability and mode of acquisition effects on choices between hedonic and utilitarian options,” *Marketing Letters,* 12 (1), 37-49.

Peterson, R. A., Albaum, G., and Beltramini, R. F. (1985). “A Meta-Analysis of Effect Sizes in Consumer Behaviour Experiments”, *Journal of Consumer Research*, 12(1): 97-103.

Schaer, O., Kourentzes, N., & Fildes, R. (2019). Estimating the market potential with pre-release buzz. Lancaster University Management School, Management Science Paper Series, (2019), 1.

Shi, Y., Karniouchina, E., & Uslay, C. (2020). (When) can social media buzz data replace traditional surveys for sales forecasting?. *Rutgers Business Review*, 5(1), 43-60.

Simon, Herbert A. 1959, “Theories of Decision-Making in Eonomics and Behavioral Science,” *American Economic Review,* 49 (3), 253-283.

Smith Lones and Peter Sørensen. 2000, “Pathological Outcomes of Observational Learning,” *Econometrica,* 68 (2), 371-398.

Steemkamp, J-B E.M., and Burgess S. M. (2002). “Optimum Stimulation Level and Explanatory Consumer Behaviour in an Emerging Consumer Market”, *International Journal of Research in Marketing*, 19: 131-150

Schauerte, R., Feiereisen, S., & Malter, A. J. (2021). What does it take to survive in a digital world? Resource-based theory and strategic change in the TV industry. Journal of Cultural Economics, 45(2), 263-293

Tversky Amos and Daniel Kahneman. 1986, “Rational Choice of the Framing of Decisions,” *Journal of Business,* 59 (4), 251-278.

Van Praag, Bernard and Ada Ferrer-i-Carbonell. 2004, “*Happiness Quantified: A Satisfaction Calculus Approach,”* London, UK: Oxford University Press.

Walls, W. David. 1998, “Increasing Returns to Information: Evidence from the Hong-Kong Movie Market,” *Applied Economic Letters,* 4 (5), 215-219.

Wasserman, Max, Satyam Mukherjee, Konner Scott, Xiao H.T. Zeng, Filippo Radicchi and Luís A. Amaral. 2015, “Correlations between user voting data budget and box office for films in the internet movie database,” *Journal of the Association for Information Science and Technology, 66* (4), 858-868.

Welch, Ivo. 1992, “Sequential Sales Learning and Cascades,” *Journal of Finance,* 47 (2), 695-732.

Weathers, Danny, Scott D. Swain and Varun Grover. 2015, “Can online product reviews be more helpful? Examining characteristics of information content by product type,” *Decision Support Systems, 79*, 12-23.

Zufryden, Fred S. 1996, “Linking advertising to box office performance of new film releases A marketing planning model,” *Journal of Advertising Research,* 36 (4), 29-41.

Appendix 1

Figure A.1:

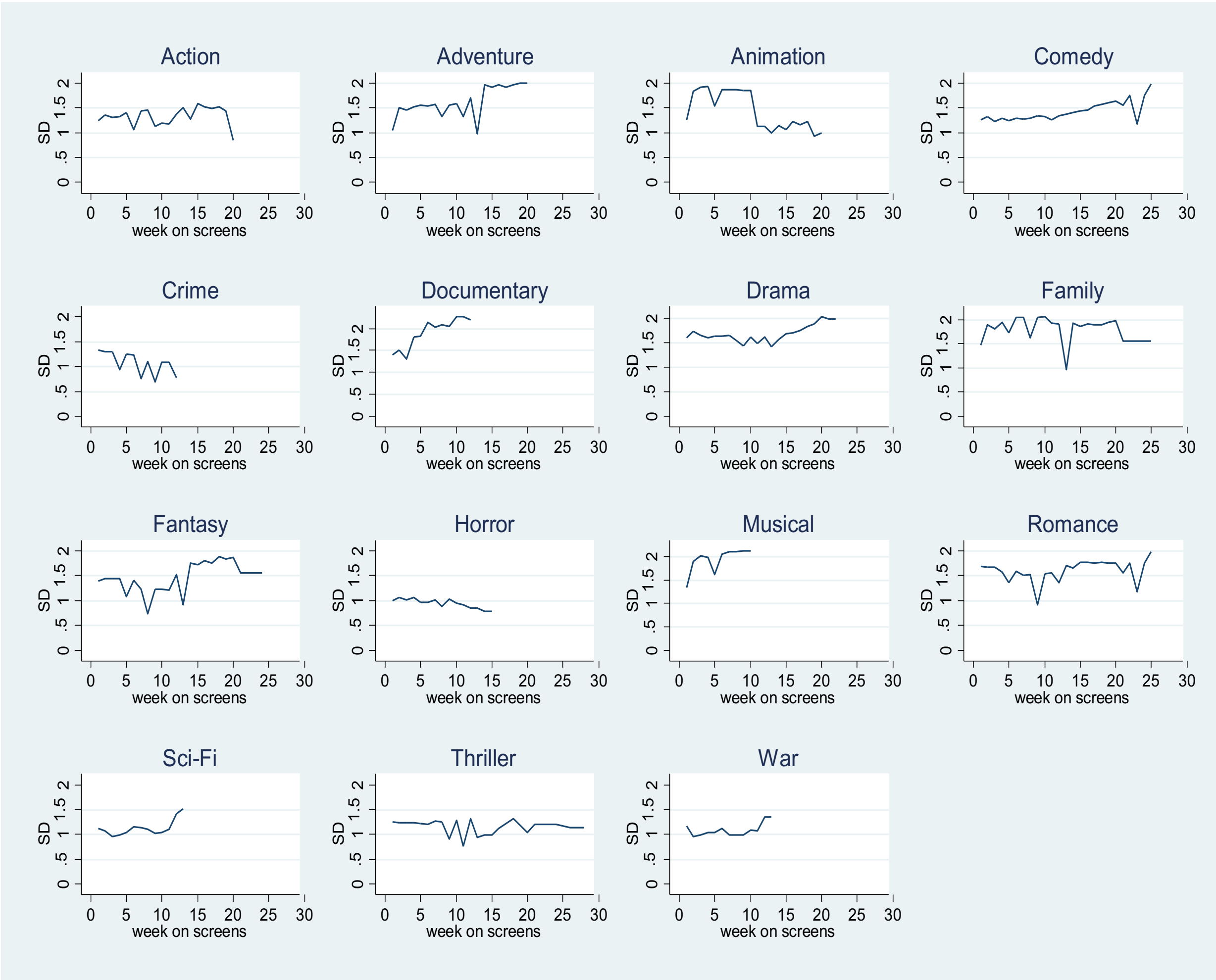
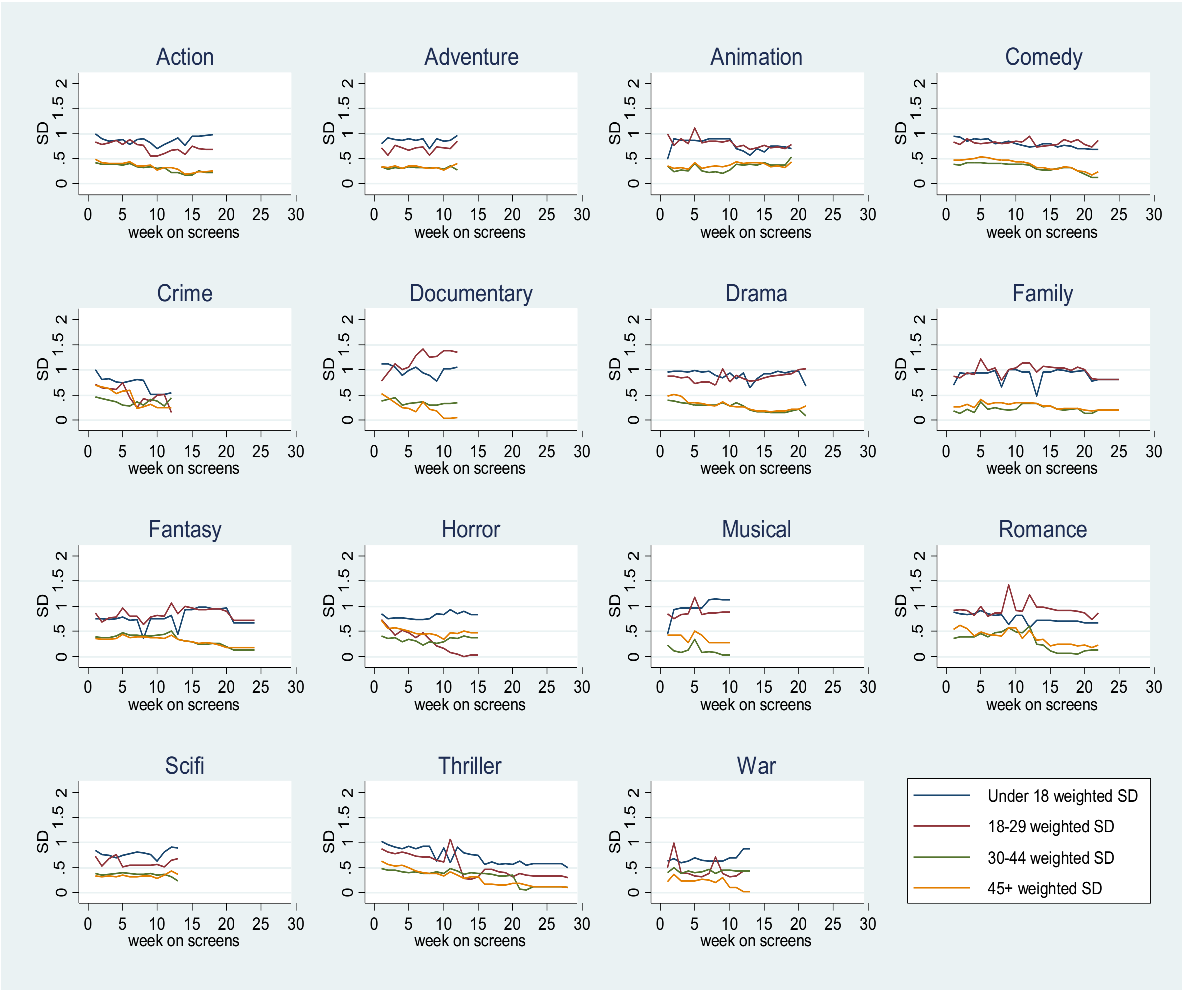


Figure A.2:


Action
Adventure
Animation
Comedy
Crime
Documentary
Drama
Family
Fantasy
Horror
Musical
Romance
Scifi
Thriller
War
SD
0
.5
1
1.5
2
0
5
10
15
20
25
30
week on screens
Under 18 weighted SD
18-29 weighted SD
30-44 weighted SD
45+ weighted SD