ANTECEDENTS AND CONSEQUENCES OF THIRD-PARTY PRODUCTS EVALUATION SYSTEMS: LESSONS FROM THE INTERNATIONAL MOTION PICTURE INDUSTRY

Mark A.A.M. Leenders
Assistant Professor of Social & Behavioral Sciences
Amsterdam School of Communications Research
Section Media Marketing
University of Amsterdam
Amsterdam, The Netherlands
Tel: 020-5256695
Fax: 020-5253681
E-mail: m.a.a.m.leenders@uva.nl

Jehoshua Eliashberg
Sebastian S. Kresge Professor of Marketing
Marketing Department
The Wharton School
University of Pennsylvania
Philadelphia, PA 19104-6371
Tel: 215-898-5246
Fax: 215-898-2534
E-mail: eliashberg@wharton.upenn.edu

October 10, 2006

We thank the various local ratings/ classifications offices for providing the ratings data as well as information concerning their structure and procedures. Mark Leenders also thanks the Erasmus Trustfund for providing financial support.

Antecedents and Consequences of Third-Party Products Evaluation Systems:

Lessons from the International Motion Picture Industry

Abstract

This paper studies one key characteristic shared by a growing number of industries—their products are continuously monitored and evaluated by third-party systems such as the Motion Picture Association of America (MPAA) and the Food and Drug Administration (FDA) for public protection. By studying differences in ratings and commercial performance of products across different countries, marketing managers can obtain general insights into why, and anticipate when evaluation boards, operating in different global markets, are likely to rate *the same* product *differently*. Since different ratings may lead to differences in commercial performance, the product evaluations often require local adaptations.

Two related issues are examined in this paper. First, the impact of the product's evaluation by the local evaluation board upon its commercial performance (consequences) is studied. Local country covariates such as mandatory vs. voluntary evaluation boards and cultural characteristics such as masculinity are employed as moderators between the product's evaluation and its local performance. Second, the key drivers of the boards evaluations (antecedents) are studied, taking into account both the nature of the product *and* the nature of the evaluation board. Implications of the results for the motion picture industry as well as other industries are discussed from a managerial, research, and public policy perspective.

1. Introduction

One of the key challenges in effectively marketing products and services in the U.S. as well as globally is the evaluation of the product's consumption by local third-party, either mandatory or voluntary system. Industries in which products are normally evaluated for public safety and protection, and where, for example, warning labels are issued, include: ethical drugs, food, motion pictures, computer games, automobiles, electrical appliances, and construction. In the U.S., for instance, the ratings classification system of the Motion Picture Association of America (MPAA) is concerned with potential harmful consequences of consumption of movies. In general, third-party evaluations may play a critical role in the success of a product. As noted by Kotler: "marketers must be aware of these regulations when proposing, developing, and launching new products" (Kotler, 2003, p.174).

Surprisingly, despite their importance and relevance, marketing researchers have given little attention to the antecedents and consequences of the products evaluation systems in any of the industries noted above. Questions of interest include: What are the key drivers of the product's evaluation? To what extent is the same product likely to be evaluated differently in various local markets around the world? Can these differences be explained and anticipated? How will these different evaluations affect the consumption of the product in the local market? We address these questions in an empirical study based on data from the motion pictures industry. However, we draw occasional parallels to other industries, in particular to the pharmaceutical industry, at different places.

The MPAA ratings system, established in the U.S. in 1968, is an example for a voluntary evaluation board for the U.S. market that evolved since 1922. It consists of a board of

about twelve women and men representing a range of racial and employment backgrounds, whose main goal is to evaluate and provide parents and moviegoers with advance information on films; thus enabling them to make judgments on movies they would like their children to watch or not to watch. The motion pictures ratings system originated in the filmmakers' community to preempt government regulation (Peacock, 2001). Its ratings are: G (for general audiences); PG (parental guidance suggested); PG-13 (parents strongly cautioned); R (restricted); NC-17 (no one 17 and under is admitted). It is not mandatory that every movie be submitted to the ratings board. However, movies that are not rated (NR) find it very difficult to receive screens and their commercial performance is minimal. From the public standpoint, the ratings board has informational value indicating what the public can expect from the movie experience. From the film marketer's standpoint, the system poses some key challenges because it may restrict the potential audience.

It has been widely observed that the commercial success of a movie may differ greatly across countries.[1] The movie *Alien Resurrection* collected $113M in box-office tickets sales outside North America and only $47M in North America. On the other hand, the movie *Ace Ventura* collected $45M outside and $108M in North America. Interestingly, such differences may be explained, in part, by the observation that the *same* movie is rated *differently* in different markets. A movie that generated quite a bit of discussion and different (restrictive) ratings around the globe is *Hannibal*. In Australia, for instance, it was rated suitable for 18+ years old (in France –12). The Australian rating imposed

[1] See www.MPAA.org for detailed descriptive statistics.

serious limitation on the studio's ability to release *Hannibal* profitably there (Australian Classifications, 22nd February 2001). [2]

The antecedents of different ratings for the same product may not only be related to the local culture, but also to the nature of the evaluation system that is in place in the particular country. For example, products evaluation systems can be voluntary (e.g., movie ratings boards in the U.S. and Germany) or mandatory (e.g., movie ratings boards in Australia, and food and drug evaluation systems around the globe). Interestingly, there is considerable variation in evaluation systems across countries and this offers interesting opportunities to study not only the effects of ratings on performance, but also the key drivers of the various ratings.

The paper is structured along the consequences and antecedents dimensions. In the first part, we examine the effect of the product's evaluation (movie's rating in a specific country) on its commercial performance (local box-office tickets sales). In the second part, we study the drivers of the local product's evaluations decisions, and propose testable hypotheses for these antecedents—conceptualized in terms of product characteristics, the third-party evaluation system's characteristics, and the country's cultural characteristics. In Section 2 we review the relevant literature and present the conceptual framework along with testable hypotheses. In Section 3 we describe the data and the analyses. Section 4 discusses the results, and Section 5 provides a summary, implications, and suggestions for further research.

[2] In the U.S., the restrictive NC-17 rating is called "box office poison" (Guardian Unlimited, July 9th 1999).

2. Relevant Literature, Conceptual Framework, and Research Hypotheses

While the effects of media on children's behavior have received attention in psychology, pediatrics and communication, the antecedents and the commercial consequences of media and entertainment products' evaluations (ratings) have received far less attention by researchers. Some exceptions include studies addressing the validity and the public knowledge of ratings (e.g., Christenson 1992, Bushman and Stack 1996, Walsh and Gentile 2001).

The MPAA ratings classification has been employed as one possible determinant of the success of movies in several descriptive and forecasting studies (e.g., Sawney and Eliashberg 1996; Ravid 1999). These studies typically find that the G, PG, and to a lesser extend PG13 ratings are associated with better performance metrics such as domestic box office tickets sales, video revenues, and return on investment. This finding may be explained by the fact that more people are allowed to watch movies with more lenient ratings. In the same spirit, De Vany and Walsh (2001) have studied whether Hollywood studios produce too many R-rated movies. They found that slates deemphasizing more leniently rated movies may be quite costly to the studios in terms of revenues, return on production costs, and profits. Neelamaghan and Chintagunta (1999) have also included the MPAA ratings as covariate in their international study. However, the results concerning the role of foreign ratings have not been elaborated upon, and the correspondence between the MPAA ratings and the varying local ratings in different countries has not been examined systematically. Bagella and Bechetti (1999), for instance, studied whether movies, restricted by the local board for an audience below the age of 18, performed worse (relative to less restrictively rated movies) at the box office in

one country, Italy. Their results suggest that there was no significant effect after controlling for the impact of stars, directors, production companies, and genre. This finding, which is relevant to our broader research question, may be explained in part by the particular national culture in which the movies were released.

We are not aware of any study that has examined systematically the consequences and antecedents of a product's evaluation, in general, and a movie's rating classification, in particular, in any discipline. The extant research on motion pictures, which has typically employed the MPAA ratings classifications in explaining and predicting success of movies in the U.S. and incidentally, albeit inappropriately, in foreign markets, has provided partial and mixed evidence. The literature in the field of communications has been concerned with the effects of violence and other hazardous media elements on the public rather than on the commercial consequences of the ratings classifications per se. Similarly, no evidence exists with regard to the antecedents of such ratings.

Employing product, evaluation board, and country characteristics, the conceptual framework used in this paper is presented in Figure 1.

Figure 1 about Here

The central construct of the framework is the product's (movie) evaluation/ rating classification. We view the rating classification as an evaluation outcome, provided by a local evaluation system (board) with certain characteristics that has the potential to affect the demand for the product. The determinants are presented on the left hand side.

2.1 The Consequences of Movies Ratings

Products evaluations, in general, and movies ratings, in particular, are communicated to people involved in the consumption of the product such as target audiences and parents. Christenson (1992), for example, tested the effect of ratings used by the recording industry in the U.S. on children attitudes. A group of middle school students gave lower evaluations to the music when the album cover had an advisory label indicating explicit lyrics, which implies appropriateness for mature audiences. Respondents in the study also reported less interest in buying this type of albums. This has been called the 'tainted fruit' effect (Christenson 1992, Bushman and Stack 1996, FTC Report 2001). In general, the tainted fruit theory posits that "warning labels should decrease the attractiveness of a given product because the product might harm the consumer" (Bushman 1998). Hence, if a media product receives a more restrictive rating, and thus there exists a tainted fruit effect, the product will be consumed by less people—only those who consider it appropriate. This is sometimes further strengthened by the fact that there may also be restrictions on the product's distribution. For example, in many countries a restricted movie cannot be screened before a certain time of the day (e.g., eight o'clock in the evening).

In contrast to the 'tainted fruit' effect, there may also be an effect that works in an opposite manner. This is in line with reactance theory (Brehm, 1966), which suggests that when a person's freedom to behave in a particular setting is threatened, s/he will experience an unpleasant motivational state that consists of pressures to re-establish the lost freedom. Consumers may respond to a restrictively rated product as if it was a 'forbidden fruit' (Klein 1993, Bushman and Stack 1996, Pechmann and Shih 1999,

Bushman and Cantor 2003). For example, such statement as “This Product is for Adult Users Only” may stimulate young people to use the product, because they seek to be more adult. When people believe their freedom is threatened, they may enter into a reactance motivational state and act to regain control by not complying with the norms. The forbidden fruit theory has been tested in different settings, including restroom graffiti, anti-smoking warnings, fatty foods warnings, drinking legislation, and media products (e.g., Pechmann and Shih 1999, Bushman and Cantor 2003). Most studies have focused on the restrictiveness of the warnings or ratings, suggesting that they are more likely to invoke forbidden fruit behavior than informational ratings or no rating at all. For movies, a youngster below the recommended age may resist the restriction and in fact be tempted to watch it. Morkes, Chen and Roberts (1997) studied the effect of different types of audience restrictions, including MPAA movies ratings, on youngsters. Respondents were first asked to read brief descriptions of movies, randomly labeled with an MPAA classification, and then to rate their attractiveness. It was found that youngsters’ desire to see the film increased as the rating restrictiveness increased.

Previous research suggests that the tainted and forbidden fruit effects of ratings classifications may occur simultaneously. The balance between these opposite effects is likely to be dependent on specific factors. Earlier research has shown that reactance state differs between genders. Brehm and Weinraub (1977) used physical barriers to impose reactance. They found that young boys preferred an object behind a barrier when the barrier was high. Girls directed their attention to the non-barricaded accessible object (more tainted fruit). Bushman and Cantor (2003) found, in particular, that boys around the age of 11 showed a tendency for forbidden fruit behavior. Swart, Ickes and

Morgenthaler (1978) found that females demonstrated reactance when the threat to freedom was low, while males showed more reactance, in the presence of a radical threat to their freedom.

The first situational factor that is of interest in our study with respect to the consequence of ratings is the local culture in the country in which the movie is released. People who live in one country are thought of as sharing a national identity and some kind of homogeneous ideology (Schlesinger 1987). Hofstede (2001) defines national culture as the collective programming of the mind and he distinguishes the masculinity, individualism, uncertainty avoidance, power distance, and time orientation dimensions. The importance of cultural dimensions in affecting consumption may differ across different industries.

In the context of motion pictures and movie ratings, the masculinity characteristic seems especially relevant. Masculinity has received much attention in historical film research in the U.S. and abroad (Powrie, Davies, and Babington 2004). Moreover, since violence is a major policy concern behind movies ratings, masculinity represents a closely associated factor. In their study on random school shootings, Kimmel and Mahler (2003) state that: "one factor that cuts across all random school shootings is masculinity." Masculinity of a country is related to gender differences, and to the duality of the sexes with which different societies cope in different ways. Men are supposed to be assertive, competitive and tough. Women--to take the tender roles (Hofstede 2001, p.280). According to Hofstede (2001), both men and women hold tougher values in masculine countries and more tender values in feminine ones. More masculine societies (e.g., Spain and Italy) tend to place greater value on wealth, success, ambition, material things, and

achievement, whereas more feminine societies (e.g., the Netherlands) tend to place greater emphasis on people, helping others, preserving the environment, and equality of life (Hofstede 2001; Lynn, Zinkhan, and Harris 1993; Steenkamp, ter Hofstede, and Wedel 1999). As a result, the expectation is that the tainted fruit effect of ratings in feminine countries will be stronger. In addition, the higher degree of reactance observed in boys compared to girls may also be linked to the effects of rating classifications. This suggests that the reactance phenomenon (i.e., more forbidden fruit) may be more prominent in countries with more masculine cultures because of values such as 'being in control' and the proneness towards aggression (Hofstede 2001, p.298-306). We therefore hypothesize that:

H1a: *The relationship between the product's evaluation and its commercial performance will be moderated by the country's masculinity. All other things being equal, the effect of a restrictive product evaluation will be less severe (negative) and may even be commercially beneficial (positive) in more masculine countries (i.e., masculine countries exhibit more forbidden fruit behavior).*

The second situational factor of interest is the regulatory system in the local country. In general, evaluation systems may be either mandatory (by the government) or voluntary (self regulation). Research on the forbidden fruit theory has paid some attention to the nature of the authority behind the products evaluations. Bushman and Stack (1996) found evidence, in an experimental setting, that warning labels for television programs increase the likelihood of forbidden fruit behavior, if the warning is perceived to come from an authoritative figure or body. This may be attributed to the increased pressure to comply,

and therefore increased pressure to exhibit reactance. Since governments are viewed as more authoritative figures than voluntary organizations, this may induce individuals to behave more in line with the 'forbidden fruit' theory, leading to a less negative effect of a restrictive rating on commercial performance of the product. Research on public health directives (e.g., crisis, smoking, or AIDS warning) has also shown that the central factor in obstructing compliance with public health messages is questions raised by consumers concerning the decision maker's authority. The more of these questions an individual has, the more s/he will be inclined to individually weigh the reasons for the action (Raz 1986; May 2004). In the popular press, it is has been often stated that warning messages attributed to a Ministry 'surely boomerang' (Guttman and Peleg 2003). Although academic research in this area is scant and has been mainly conducted in the context of AIDS and smoking, health messages disseminated by the government are known to score low on public believability (Guttman, Boccher-Lattimore, and Salmon 1998). We hypothesize:

H1b: *The relationship between the product's evaluation and its commercial performance will be moderated by the regulatory status of the evaluation system. All other things being equal, the effect of a restrictive evaluation will be less severe (negative) in countries where the product's evaluation system is mandatory (i.e., mandatory evaluation systems lead to more forbidden fruit behavior).*

2.2 The Antecedents of Ratings Classifications

In addition to investigating the *consequences* of evaluations made by third-party products evaluation systems on commercial performance, it is imperative for marketers to

understand the *drivers* of these evaluations. We identify two key drivers: product characteristics and the evaluation board characteristics.

Characteristics of the Product: The Case of Movie's Mechanisms

Movies mechanisms used for creating entertainment experiences as well as for artistic/ aesthetic expression that may be hazardous to the public, are labeled by the MPAA in terms of 'violence', 'sex', nudity, and 'language'. [3] In terms of side effects, these mechanisms are elements that may cause potential harm or disruption to certain individuals such as youngsters who are looking at the world from a distorted perspective and may imitate the "real world" behavior they see in the movie. Moreover, violence and possibly aggressive behavior sometimes occur because media violence increases physiological arousal and thereby intensifies subsequent emotional responses. Harmful mechanisms such as violence are often augmented by other potentially less harmful mechanisms such as 'rough humor', 'gore', and even 'sensuality'. All are essential for the creative process. Once the creative process is completed, the evaluation process typically works in such a way that members of the ratings board view the movie, discuss it, and vote on the film's rating. If the movie's distributor is not satisfied with this rating, s/he can

[3] We use the term mechanism here for two reasons. First, in line with social psychology, these detailed product characteristics (more than general characteristics such as genre) are the "generative mechanisms" that influence the dependent variable of interest here: the rating (e.g., Baron and Kenny 1986). Second, in ethical drugs, a product category in which the conceptualization is also applicable, the detailed characteristic of the medicine that drives the effect is called the mechanism of action. For example, the mechanism of action of the painkiller VIOXX is the inhibition of prostaglandin synthesis. Its key efficacy is that it is a powerful painkiller that is well tolerated by the stomach. However, one of its key side effects is that it slightly increases heart risks.

re-edit the film and re-submit it, but at the same time, has to face very strong resistance from the creative community (director, actors, and actresses) and (often) a costly launch delay. From all the mechanisms noted above, exposure to violence and its effects on society received much attention and publicity. It is estimated that the average child watches 21-23 hours of television per week, and from cartoon programming only, may watch 200 violent incidents, along with 16,000 simulated murders and 200,000 other violent acts (Caron 2001; Federal Communications Committee 2003). School shootings and random violence have been in the news around the globe (Grier 2001). In particular, the U.S. ranks first among all industrialized nations in violent death rates, and the Department of Health and Human Services has viewed "Violent and Abusive Behavior" as one of its top priority areas since the early 90's. So, although members of the entertainment industry sometimes argue that the relationship between watching media violence and aggressive behavior is bi-directional (e.g., Black and Bevan 1992), the importance of violence for rating decisions has been widely established. Hence, we hypothesize that of all mechanisms employed to create a movie, the violence mechanism is the most prominent driver of restrictive ratings worldwide:

H2a: *All other things being equal, the violence mechanism in movies is the primary driver of restrictive ratings.*

Although we expect that the use of the violence mechanism is the major driver of restrictive ratings, given the cultural differences between countries, we also expect to observe differences across countries in how the boards react to this mechanism. In masculine cultures, where children learn to admire characters such as *Batman* and

Rambo, the reaction may not be as stiff as in feminine countries where the 'role model' is *Ollie B. Bommel*, the clumsy anti-hero (Hofstede 2001). We thus hypothesize:

H2b: *The relationship between the violence mechanism in movies and their ratings will be moderated by the country's cultural masculinity. All other things being equal, the violence mechanism will lead to more restrictive ratings in less masculine countries.*

Characteristics of the Evaluation Boards

Movies ratings are determined by people, typically organized as a board. After reviewing the film, the board members follow a deliberation process that ultimately leads to the final rating. Many evaluation systems collect periodically feedback from their publics (e.g., complaints, surveys). Three major characteristics of the evaluation board are: its composition in terms of experts versus lay people (novices), the extent of industry involvement, and its size. Risk Assessment is also a critical aspect of the process.

The differences in risk assessment between experts and lay people have been studied extensively (e.g., Fischhoff, et. al 1981). It has been shown that experts do not necessarily assess risks better than lay people, even when both are presented with the same available data. One reason for this is that many risk problems force experts to go beyond the limits of the available data and convert their incomplete knowledge into judgments by falling back on intuitive processes and rule of thumb judgments much like those employed by lay people. This is a possible reason why lay people may even be present in science-driven evaluation boards that advise the FDA on new medicines (FDA consumer

magazine, September 2000). In some FDA committees, there are lay people who are "representatives of consumer or patient interests".[4]

Previous research has also shown that lay people tend to be more cognitively involved in issues related both to their self-interests as well as to the interests of others for whom they have empathy. As personal interest in some topic or event increases, so does issue-relevant thought. People work hard to form opinions on topics that portend positive or negative outcome relevant to their interest (Petty & Cacioppo, 1981, 1986). As a result, cognitively involved people are likely to perceive more types of risks and even frame them more as losses (Kahneman and Tversky, 1973). We expect to see similar behavior in the case of movies ratings, where often parents make up part of the board. In addition, people highly involved in an issue see *all* topic-related arguments as being important (Heath & Douglas, 1990). Karger and Wiedemann (1997) environmental risk assessment study show that: "Lay people have a 'worst-case' scenario in mind. All environmental risks were appraised as high risks. This applies to all types of environments such as air pollution, water pollution, and contamination of soil. Furthermore, lay people are quite sure of their assessment."

Finally, evidence suggests that in criminal justice, lay people when acting as jurors, tend to respond emotionally to various facts and consequently impose stiffer punishments. Prentice (2003) reviewed studies related to behavioral analysis of law that have shown that "mock jurors will tend to impose stiffer punishments against muggers who attack a person on her way home and to return higher damage awards against careless drivers who hit a person on her way home if the victim is taking an unusual route home. In an unusual

[4] FDA guidelines. Advisory Committees: Implementing section 120 of the Food and Drug Administration Modernization Act 1997.

setting, it is easy for jurors to imagine 'if only she had been taking her usual route home, this wouldn't have happened.' This counterfactual thinking increases the jurors' emotional response to the facts of the case and, in turn, affects their judgments. In fact, in criminal justice, professional judges tend to be more lenient in the court than lay people acting as jurors (Institute for Security Studies, 2000). Based on the above, we hypothesize that:

H3a: *All other things being equal, the presence of lay people in the evaluation board will lead to more restrictive ratings*.

The second characteristic that may drive differences in ratings is industry involvement in the evaluation board. The role of the industry in movies evaluations is a recurring theme in the popular press worldwide. Interestingly, in film and in other industries, there is considerable variation in the degree of industry involvement around the globe.[5] In the motion picture industry, there are countries with no industry involvement in the board (e.g., Australia, Hong Kong, and Italy). In the U.S. and U.K., there is some industry involvement because the ratings are provided by industry-backed organizations such as the MPAA and BBFC, respectively. In Germany, some members of the ratings board are appointed by the industry. Recently, the Netherlands have installed an experimental system where the movies' distributors are completely responsible for rating their own movies, based on guidelines provided by the classification office *Kijkwijzer*. Since economic considerations drive most industries, including movies, industry involvement is

[5] Even in the pharmaceutical industry, indirect industry involvement cannot be denied. The recent revote on Vioxx and other similar medicines - based on additional evidence of heart risk after longer use - showed that 10 people from the voting committee had previous consulting relationships with the drug's makers. The 10 advisers with company ties voted 9-1 to keep Vioxx on the market (www.fda.gov).

likely to lead to more lenient evaluations that do not 'taint the fruit' and thus exclude too many consumers.[6]

H3b: *All other things being equal, industry involvement in the evaluation board will lead to more lenient evaluations.*

The third characteristic, over which ratings boards often differ, is the size of the board and its effect on the group decision-making. One phenomenon that has been studied extensively in the context of small group decision-making is group polarization, which posits that group discussion has a tendency to enhance individuals' positions on issues in one of two possible directions: risky shift or cautious shift (Isenberg 1986). The direction depends on the specific problem the group is facing. Stoner (1961) observed that groups' decisions were riskier than the decisions made by the individuals comprising the group. The observation has later been explained by a number of social norms (Myers and Lamm 1976; Crott and Zuber, 1983). Sunstein (2000) provides examples of problems that produce cautious shift. Various studies have found that as the size of a group becomes larger (e.g., from five to twelve members) the degree of consensus decreases (Hare 1952, Devine et al. 2000) and the group members' conformity to normative group pressures increases (Kessler 1973; Saks 1977). Hence, larger groups are more likely to behave in a socially desirable manner (Kohli 1989; Ridgeway 1983). Since rating boards are expected *a-priori* to be cautious and less risky in their behavior, we hypothesize that larger boards will behave more cautiously, and hence, make more restrictive decisions.

[6] That does not mean that studios can or will always press for the 'General Admission' rating. The movie industry recognizes the information importance of the ratings. There are cases (Harry Potter 3) where the movie distributor accepts a newly created rating classification (older than nine) in some countries (The Netherlands) as a compromise (Goldsmith 2004).

H3c: *All other things being equal, the larger the size of the evaluation board, the more restrictive will be its evaluations.*

3. Data, Sources, Variables, and Analysis Approach

3.1 The Data

Data were collected on all internationally released mainstream movies distributed by the major studios in the U.S. between December 29th 1996 and January 3rd 1999. Data on this set of movies were also collected in the following markets: Australia (Aus), Hong Kong (Hon), Italy (It), Spain (Spa), Germany (Ger), France (Fra), and the U.K. (UK). The movies included in the data set represent U.S. productions, U.S. and other country co-productions, and pure non-U.S. (foreign) productions. The eight countries studied represent the top ten markets with respect to box-office potential. The box-office performance of the movies in each country, measured in dollars for the opening weekend, was compiled from *Showbizdata.com* and *VNU/ ACNielsen* databases. We focus on the opening weekend box-office performance in order to eliminate the impact of various time-dependent variables (e.g., word-of-mouth) and because it is highly correlated with the cumulative box office performance (see, for example, Sawhney and Eliashberg 1996). Our other data sources are discussed below.

The Dependent Variable:

In order to adequately compare the opening weekend box-office across different countries, it was normalized, by dividing the opening weekend box-office by the peek weekend ticket sales (in dollars) for the complete top 10 in the country during the data collection period, excluding the tickets sales from purely local movies. This variable,

called demand share (DSH), reduces the need for a broad range of country-related controls in the analyses. The demand share variable is given by:

$$DSH_{ij} = \frac{\text{Opening weekend box office receipts of movie i in country j}}{\text{Maximum box office weekend receipts (top 10) in country j during data collection}} \quad (1)$$

The Independent Variables:

Screens

The number of screens on which a movie plays in its opening weekend has been shown to be a significant variable in explaining movies performance (Litman 1983, Elberse and Eliashberg 2003). Data on the number of screens in the different countries were also compiled from Showbizdata.com and VNU/ ACNielsen. The number of screens during the opening weekend was normalized in a fashion similar to the demand share variable. That is, the number of screens allocated to the movie during its opening weekend in a country was divided by the maximum total number of screens allocated to the top-10 movies for any weekend during the data collection period (excluding screens allocated to local movies). This variable is called screen share (SSH).

Ratings Classifications

Ratings classifications (RAT) of the movies released theatrically in the different markets were obtained from the *local* rating and classification office in the particular country that was studied. Special attention was given to ensure that it was the rating of the theatrical release, as the ratings of the trailer, the video, or DVD often differ. Since most ratings classifications are based on an age cut-off point, the movies ratings in the different countries were also recoded on a common numerical scale, corresponding to the

minimum age that is recommended to watch the movie (without being accompanied by parents). This coding scheme (see Appendix 1) allowed us to make a comparison across countries and to allow for a broader set of analyses techniques, where a lower score indicates a less restrictive board classification. For the PG rating, used in the U.S. and a few other countries, the ratings boards provide no clear age cut-off point since the rating only means 'parental guidance' or 'not suitable for 'children'. Based on telephone interviews with the local boards and their suggestions, the PG rating was given the cut-off point of 10 years old and therefore a score of 10. In addition to the ratio scale version, an unrecorded ordinal scale was used.

Movie Mechanisms

Note that H2a has been formulated in terms of the impact of violence relative to other mechanisms (e.g., language, sex, nudity). These mechanisms were compiled from the MPAA database, www.MPAA.org. They were coded as dummy variables where 1 denotes the presence of the mechanism and 0 its absence (one movie can contain multiple mechanisms). The MPAA considers a wide range of mechanisms, not all of them are stated in each movie and in our study. The mechanism was included if it appeared in at least 5 movies. The mechanisms studied and the number of movies containing them (in parentheses) was: Language (150), Violence (93), Sex (66), Thematic Elements (25), Nudity (20), Sensuality (18), Drugs (16), Rough Humor (16), and Gore (10).

Board Characteristics

The composition of the evaluation board was obtained from documentation available at the local ratings boards' offices, and from local guidelines or statutes regarding the formation of the board in each country. In particular, the average size (SIZ) of the board

was measured directly by the average number of people serving on the board; a dummy variable was used to indicate whether there are experts (EXP) in the board (EXP = 0 if not included, EXP = 1 otherwise); another dummy variable was used for whether the board is mandatory (MAN) or voluntary (MAN = 0 if voluntary, MAN = 1 if mandatory). The board was considered a mandatory system if the evaluation board is part of a ministry, or if the minister itself is responsible for the ratings. If the board is a private/ industry organization, it was considered as a voluntary system. Finally, industry involvement in the board (IND) was measured such that IND = 0 if none, IND = 1 if industry involvement is only in the appeal process, IND = 2 if industry can select the board's members, IND = 3 if industry representatives are present in the evaluation board.

Country Characteristic

The countries Masculinity (MAS) indices were taken from Hofstede (2001). Scores from a validation study conducted by Hoppe (1998) were used to validate the results. The scales range from 0 to 100, with higher values indicating higher degree of masculinity. Table 1 provides key descriptive statistics.

Table 1 about Here

The table shows that the number of movies in the data set during data collection was largest in the U.S. This is due to the fact that for some movies no box office data were available outside the U.S. In addition, some data were missing in some countries because of incomplete ratings data or because it was not clear whether or not the rating was for the theatrical release.

Table 1 also shows some interesting differences and similarities across countries. First, the average ratings are such that they tend to be more lenient in France and Italy and most

restrictive in the U.S., Australia, and the U.K. The demand shares are quite comparable across the countries, although it is a somewhat lower in Italy consistent with somewhat lower screen share in that country. This, perhaps, reflects the Italian distributors/ exhibitors' tendency to avoid a 'blitz' distribution strategy, which draws smaller audience.

3.2 Analysis Approach

To test the hypotheses concerning the consequences of the ratings classifications, a logistic regression model accommodating random movie and country effects was used. The logistic form can accommodate a flexible S-shaped function as well as a market saturation level. The S-shaped form captures the essential impact of the screen share. It is consistent with other studies indicating that product availability captured by shelf space and booth size is related non-linearly to sales (e.g., Gopalakrishna and Lilien 1995, Dekimpe et al. 1997). Since the dependent variable DSH ranges between 0 and 1 (the saturation level is 1), the logistic form can be written as:

$$Log\left(\frac{DSH}{1-DSH}\right) = \alpha + \sum_{j=1}^{j} \beta_j x_j \qquad (2)$$

Note that equation (2) can also include interaction terms, needed to test the moderation hypotheses (Baron and Kenny 1986). A mixed regression estimation procedure was used to accommodate two random effects (δ_i and γ_j in addition to ε_{ij}) as there is a possibility that the observations are dependent within movie i and within country j because the same movie is observed several times in eight countries (see Greene 2002). To test the antecedent's hypotheses, two analytical approaches have been employed for robustness

testing: mixed linear regression and Ordered Probit. To reduce multicollinearity, the variables in the interaction terms were mean centered (Jaccard et al. 1990, Cohen and Cohen, 1983).

4. Results

4.1 The Consequences of Movies Ratings

The pair-wise correlations between the main variables are presented in Table 2. The demand share variable has the highest correlation ($r = .67$, $p < .01$) with the screen share. This is in line with earlier research (Elberse and Eliashberg 2003). The correlation between the demand share and the rating classification is negative and significant ($r = -.249$, $p < .01$). The negative sign indicates that, without controlling for other variables, more restrictive ratings are associated with a lower commercial performance of movies during their opening weekend.

Table 2 and 3 about Here

The results of the logistic mixed regression specified in equation (2) are presented in Table 3. The table also shows that, as a main effect, the rating variable (RAT) obtains a significant negative coefficient. However, since interactions between the RAT variable and other variables are included in the model, this main effect needs to be interpreted cautiously: It represents an "average effect across the different values of the moderator" (see, for example, Jaccard et al. 1990, p.14). In this context, this finding can be interpreted such that after controlling for the other variables and the random effects, there is a general tainted fruit tendency, which is in line with earlier research that has used the MPAA rating as a control for explaining box office performance.

The interaction between the rating classification and the country's masculinity is significant ($p < .05$, $t = 2.60$)[7]. Figure 2 demonstrates that there is indeed a stronger forbidden fruit effect in the more masculine countries compared to the more feminine countries. Similar results were found when the cumulative box office revenues were used as dependent variable, albeit somewhat weaker, and with the masculinity index of Hoppe (1998). Hypothesis H1a is thus supported.

The interaction RAT*MAN is not significant. Therefore, we find no support for the argument that governments trigger more forbidden behavior than voluntary organization. Possibly the potentially higher forbidden-fruit effect is mitigated by stronger enforcement policies at the theatre door in regulatory environments. In sum, the empirical results do not lend support for hypothesis H1b. [8]

Figure 2 about Here

4.2 Different Ratings Classifications for the Same Product in Different Countries

A question of interest arises concerning the extent to which differences in rating classifications of the *same* movie exist in *different* countries. To examine this question more closely, the sample of movies shown in the U.S. and various other countries is matched, using the U.S. ratings as benchmark in each pair of countries. Table 4 illustrates the results.

Table 4 about Here

[7] The decrease in –2LL (smaller is better) related to Mas * Rat is significant $p<.05$.

[8] Since there are different classification categories across different countries, the interaction results may be somewhat magnified or mitigated. To validate our findings, the RAT scale was transformed into a dummy variable coded as 0 if the movie was allowed for all ages (a classification that is available in all countries), and as 1 if that was not the case. The results were quite similar.

Recall that a lower rating score implies a more lenient evaluation. The table shows that the differences in ratings of the same movie, between the U.S. and the other counties, are significant such that *all* Non-U.S. countries tend to be *more* lenient, relative to the U.S., with the largest differences occurring between the U.S. and France (14.04 vs.1.55), the U.S. and Italy (14.54 vs. 2.69), and the U.S. and Spain (13.87 vs. 7.70). This finding is very much in line with a recent study conducted on behalf of the European Commission (Olsberg/SPI and KEA 2003). The last three columns of Table 4 show that similar differences exist after the sample of matched movies is reduced further to include only those movies that have runtimes in the Non-U.S. market identical to that in the U.S., thus controlling better for their content.

4.3 The Antecedents of Ratings Classifications

Table 5 presents the estimation results (mixed linear and Ordered Probit) needed to test H2a-H3c. The independent variables are categorized in movie mechanisms, culture and board characteristics.

Table 5 about Here

Of all the mechanisms, violence shows the strongest positive effect on ratings in both the mixed linear regression and the Ordered Probit models. In addition, the differences between the violence coefficient (B = 3.11) and the other mechanisms are significant ($p < .01$) except for the difference with the 'Gore' coefficient (B = 2.92). Hence, overall, we find empirical support for H2a. Interestingly, we find a significant negative effect for 'Sensuality'. This suggests that 'Sensuality' may be viewed by ratings boards in some countries as beneficial for younger viewers. This finding raises the question why more beneficial mechanisms (e.g., educational and artistic mechanisms) are not included in the

ratings criteria. In other industries such as the pharmaceutical industry, the benefits (efficacy) of the product are always weighed against the cost (side effects).

The interaction between the country's masculinity and the violence mechanism is significant in both models (albeit for the mixed linear regression model only at $p < .1$ (one-tailed)). The increase in explained variance due to the interaction is significant in the Ordered Probit model, but just not in the linear regression model ($p = .14$). Dividing the sample into masculine and feminine sub-samples, indicates that the nature of the interaction is in line with the hypothesis: more masculine countries tend to be more lenient towards the violence mechanism. H2b is thus weakly supported.

The empirical results concerning the effects of the board characteristics on the rating classifications are generally in line with the hypotheses. The fit statistics for the separate categories of independent variables indicate that board characteristics explain most of the variance. The inclusion of experts in the board is associated, as hypothesized, with more lenient ratings ($p < .001$). Larger boards are indeed related to more restrictive ratings. Industry involvement, however, is not necessarily associated with less restrictive ratings classifications. In sum, there is strong support for H3a and considerable support for H3c, but not for H3b.

5. Summary and Discussion

In this study, we developed and tested a framework for the antecedents and consequences of third party product evaluations in the context of international motion pictures. Our results show that movies ratings play a significant role in determining the movie's commercial success in the sense that they can exclude consumers (tainted fruit) as well as

attract consumers (forbidden fruit). The balance between these competing forces is to some extent dependent on the local culture and our data suggest that restrictive ratings have less negative effect on the commercial success of movies in more masculine countries. We note that in other industries and settings, different cultural dimensions may be more relevant as research on car driving, speed limits (classification system), and traffic deaths has shown that uncertainty avoidance is the primary factor in explaining the speed limit in the country as well as the number of traffic deaths (Hofstede 2001, p.199).

Our results indicate that ratings matter and that some of their drivers are identifiable and measurable. In particular, not only the content is responsible for the ratings but – and this has not been studied before - the characteristics of the ratings organization such as the size of the board and the involvement of experts versus lay people are important predictors of how lenient or restrictive the ratings are in a certain country.

Implications for Marketing Managers

Product developers and markets, working in industries where products are evaluated by third-party evaluation systems, should recognize the tradeoff between the product's appeal and its potential evaluation. Since the demand is sensitive to these evaluations, the appropriate international marketing strategy may "local adaptation". In the motion picture industry, there are already different versions of DVDs in place with different directors' cuts as well as movies with different scenes and endings.

Another implication from our study for marketing relates to the question of whether or not to appeal the local board's evaluation. As far as finished products are concerned, marketing managers should be aware that appealing procedures can be lengthy and costly and thereby they should be considered only in situations where the tainted fruit effect is

likely to be dominant, at least from a pure commercial standpoint. Of course, the moral and ethical implications of the forbidden fruit consumptions should also be considered seriously. In addition, ratings classifications for media and entertainment products can be technologically enforced with innovations like the V-chip for television sets and restrictive search bots for children surfing the Internet. Software packages that delete certain images and language from DVD's while they are playing are also available. However, the potential reactance behaviors that result from these measures are ill understood.

Implications beyond the Motion Picture Industry

We believe that our study has broader implications for industries other than the motion pictures. The MPAA evaluation system is akin to other (media) products evaluation systems for different types of vulnerable consumers. For example, it is also similar to the FDA's use-in-pregnancy ratings where medical products rated as A indicate that controlled studies show no risk; B rating indicates no evidence of risk in humans (despite some adverse findings in animals); C rating indicates risk cannot be ruled out; D rating indicates positive evidence of risk; and X rating indicates contraindicated pregnancy (i.e., risk clearly outweighs any possible benefit to the patient). In fact, organizations such as the FDA can request more or less restrictive labeling for any ethical drug. Along the same lines, the Food and Nutrition Information Center (FNIC), which is part of the U.S. Department of Agriculture, provides the Healthy Eating Index (HEI). The HEI is a score [0-100] indicating to consumers the overall quality of their daily diet. It is comprised of total fat, saturated fat, cholesterol and sodium intakes.[9]

[9] Food and drug evaluations are older than the evaluation systems employed in the media industry. From the beginning of civilization, people have been concerned regarding the quality and safety of food

A key reason why product ratings are important for marketing in a range of industries is because products evaluations potentially affect demand. For example, the average effect of more restrictive ratings for mainstream movies across the countries is negative in our study (tainted fruit is more dominant). However, there may be specific countries (as we found) or other industries with certain characteristics where a product evaluation has a different effect. Tainted (desired) and forbidden (undesired) behaviors are likely to occur in any industry. For example, the product Epogen©, developed by the pharmaceutical company Amgen, was launched for some kidney indications. However, as the product affects the red blood cell production, the product proved also very suitable for consumers engaged in sports activities. This has led to a very strong surge in (forbidden fruit) demand. Another example is the cardio-vascular market in which the class of ethical drugs called beta-blockers is also named as "the musician's underground drug", because they reduce anxiety and hand trembling.

Understanding the antecedents of products ratings also has implications beyond the motion picture industry. In the pharmaceutical industry, for instance, there are cases where a drug was rejected in one country and approved in another. Miglustat©, the Gaucher disease therapy was rejected by the FDA in 2002, which demanded further clinical studies to prove its safety and efficacy. A month later, the corresponding European board, the Committee for Proprietary Medicinal Products (CPMP), recommended the drug for approval (Nature, September 2002).[10] In addition, the size of the board differs across countries and products in the pharmaceutical industry. For

---

and medicine. In 1202, King John of England proclaimed the first English food law, the Assize of Bread, which prohibited adulteration of bread with such ingredients as ground peas or beans.

[10] The FDA's rejection followed from neuropathy symptoms shown by some patients in clinical trials, which some say are associated with the disease, as opposed to the medicine.

example, the size of the committee that evaluated Vioxx was 32, whereas a recent revote on silicone breast implants had a committee of 9 advisors. In October 2003, the breast implant committee consisted of 15 members.

From a public policy perspective, our results provide interesting insights into a key question linked to rating systems in general: Is the board too restrictive or too lenient? Given that the objective of these evaluation systems is to minimize health and safety hazards, the movies-related empirical results imply that the structure of board can become a policy instrument for establishing more or less restrictive evaluations. More specifically, if the goal is to have a more restrictive board, the results suggest that it should be large, have less industry involvement, and it should be comprised of lay people that represent the general public. Whether or not the goal should be to have more or less restrictive board decisions depends, of course, on the actual public health and safety situation in the country. Hence, the situation should be monitored on a continuous basis and the appropriateness of the board be reviewed periodically.

Limitations and Future Research

While our hypotheses were stated quite generally, they have been tested only with aggregate data from the motion picture industry. The use of segment-level data (e.g., older vs. younger than seventeen year old) would allow answering interesting questions such as: does an "R" rating lead to different effects in different segments? This question is important, and appropriate data are likely to become available, if more movies exhibitors will follow GKC Theatres' recent policy of allowing non-accompanied teenagers to watch R-rated movies by showing a parent approved pass card (Dennis

2004). In addition, measuring tainted and forbidden behaviors more directly would shed important insights.

Testing the hypotheses with data from other industries, where products evaluation systems play a major role provides an extremely important avenue for future research. This would enhance the 'highlighting' of cross-industry commonalities, and hence, researchers ability to properly design cross-industry studies. Also, the degree to which these third-party evaluations ultimately play out is related to additional factors, including the awareness of these ratings among the public. In our empirical setting of movies, the knowledge of these ratings tends to be very relatively high.

Finally, motion pictures are 'cultural' products that contain mechanisms to which the evaluation system responds. Since many ratings systems' mission statements highlight the fact that it is their aim to 'reflect' society rather than to lead it, the boards' evaluations can be viewed as providing valuable cultural information such as how various countries/cultures feel about sex, violence, and drugs over time. In our study, for instance, we found that three countries with the least lenient ratings are all of British heritage (U.S., UK, and Australia). Such information can be used to guide future studies on cultural dimensions that may be relevant in other industries.

References

Bagella, Michele. and Leonardo Becchetti (1999), "The Determinants of Motion Picture Box Office Performance: Evidence from Movies Produced in Italy," *Journal of Cultural Economics*, 23 (4), 237-56.

Baron, Ruben M., & Kenny, David A. (1986), "The Moderator-mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." *Journal of Personality and Social Psychology*, 51(6), 1173-1182.

Black, Stephen L. and Susan Bevan (1992), "At the Movies with Buss and Durkee: A Natural Experiment on Film Violence," *Aggressive Behavior*, 18, 37-45.

Brehm, Jack W. (1966), *A Theory of Psychological Reactance*. New York: Academic Press.

Brehm, Sharon and Marsha Weinraub (1977), "Oppositional Behavior in Two-Year-Old Children: A Reactance Theory Interpretation," *Journal of Personality and Social Psychology*, 35, 830-36.

Bushman, Brad J. (1998), "Effect of Warning and Information Labels on Consumption of Full-Fat, Reduced-Fat, and No-Fat Products" Journal of Applied Psychology, 83 (1), 97-101.

Bushman, Brad J. and Angela D. Stack (1996), "Forbidden Fruit Versus Tainted Fruit: Effects of Warning Labels for an Attraction to Television Violence," *Journal of Experimental Psychology: Applied*, 2 (3), 207-26.

Bushman, Brad J. and Joanne Cantor (2003), "Media Ratings for Violence and Sex, Implications for Policy Makers and Parents," *American Psychologist*, 58 (2), 130-141.

Caron, Emily R. (2001), "Blood, Guts, and First Amendment: Regulating Violence in the Entertainment Media", The Kansas Journal of Law and Public Policy, Fall 11(89).

Christenson, Peter G. (1992), "The Effects of Parental Advisory Labels on Adolescent Music Preferences," *Journal of Communications*, 42 (1), 106-13.

Cohen, Jacob, & Cohen, Patricia (1983). *Applied Multiple Regression/correlation Analysis for the Behavioral Sciences* (2nd edition). Hillsdale, NJ: Erlbaum.

Crott, Helmut W. and Johannes A. Zuber (1983), "Biases in Group Decision Making," In *Decision Making under Uncertainty*, Roland W. Scholtz ed. Amsterdam: Elsevier Science Publishers.

Dennis, Jan (2004), "Movie Chain Lets Parents Give Under 17s a Pass-to R-rated fare" *Philadelphia Inquirer*, June 13.

De Vany, Arthur and David Walls (2001), "Does Hollywood Make Too Many R-rated Movies? Risk, Stochastic Dominance, and the Illusion of Expectation," *Working Paper*, Department of Economics, University of California at Irvine.

Dekimpe, Marnik G., Pierre François, Srinath Gopalakrishna, Gary L. Lilien, and Christophe Van den Bulte (1997), "Generalizing About Trade Show Effectiveness: A Cross-National Comparison," *Journal of Marketing* 61 (October): 55-64.

Devine, Dennis J., Laura D. Clayton, Benjamin B.Dunford, Rasmy Seying, and Jennifer Pryce (2000), "Jury Decision Making: 45 Years of Empirical Research on Deliberating Groups," *Psychology, Public Policy, and Law*, 7, (March), 622-727.

Elberse, Anita and Jehoshua Eliashberg (2003), "Demand and Supply Dynamics for Sequentially Released Products in International Markets: The Case of Motion Pictures," *Marketing Science*, 22 (3), Summer, 329-354.

Federal Communications Committee (2003), "Comments on Media Policy Coalition" April 21.

Fischhoff, Baruch, Sarah Lichtenstein, Paul Slovic, S. L. Derby and R. L. Keeney (1981), *Acceptable Risk*. New York: Cambridge University Press.

Greene, William H. (2002*), Econometric Analysis*, 5th Edition, Prentice Hall, NY.

Grier, Sonya A. (2001), "The Federal Trade Commission's Report on the Marketing of Violent Entertainment to Youths: Developing Policy Tuned Research," *Journal of Public Policy and Marketing*, 20 (1), 123-32.

Goldsmith, Charles (2004), "Harry Potter Film Draws Stiffer European Ratings," *The Wall Street Journal*, June 1, B7.

Gopalakrishna, Shrinath and Gary Lilien (1995), "A Three-stage Model of Industrial Trade Show Performance," *Marketing Science*, 14(1), 22-42.

Guttman, Nurit, Daria Boccher-Lattimore, and Charles T. Salmon (1998), "Credibility of Information from Official Sources on HIV/AIDS Transmission," *Public Health Reports*, 113, 165-471.

Guttman, Nurit and Hannah Peleg (2003), "Public Preferences for an Attribution to Government or to Medical Research Versus Unattributed Messages in Cigarette Warning Labels," *Health Communication*, 15(1), 1-25.

Hare, Paul A. (1952), "Interaction and Consensus in Different sized Groups," *American Sociological Review*, 17, 261-267.

Heath, Robert L., & Douglas, William (1990), "Effects of Involvement on Reactions to Sources of Messages and to Message Clusters." In J. E. Grunig & L. R. Grunig (Ed.), *Public Relations Research Annual* (pp. 179-183). Hillsdale, NJ: Lawrence Erlbaum Publications.

Hofstede, Geert H. (2001), *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations across Nations*. Thousand Oaks, California: Sage Publications.

Hoppe, Michael H. (1998), "Validating the Masculinity/ Femininity Dimension on Elites from Nineteen Countries." In G. Hofstede and Associates, *Masculinity and Femininity: The Taboo Dimension of National Cultures* (p.29-43), Thousand Oaks, CA, Sage Publishers.

Institute for Security Studies (2000), "Justice versus Retribution," *Monograph 45*, www.iss.co.za.

Isenberg, Daniel, G. (1986),"Group Polarization: A Critical Review and Meta-Analysis" *Journal of Personality & Social Psychology*, 50, 1141.

Jaccard, James, Robbert Turissi, and Choi K. Wan (1990), *Interaction Effects in Multiple Regression*, Sage Publications.

Kahneman, Daniel and Amos Tversky (1973): "On the Psychology of Prediction," *Psychological Review*, 80, 237-51.

Karger, Cornelia R. and Peter M Wiedemann (1997): "Perception and Appraisal of Environmental Problems: Lay People's Worst-Case Assumptions," *Society for Risk Analysis Annual Meeting*.

Kessler, Joan (1973), "An Empirical Study of Six- and Twelve-Member Jury Decision-Making Processes," *Journal of Law Reform*, 6, 712-34.

Kimmel, Michael S. and Matthew Mahler (2003), "Adolescent Masculinity, Homophobia, and Violence, *American Behavioral Scientist*," 45(10), 1438-1458.

Klein, Richard (1993), *Cigarettes are Sublime*, Durham, NC. Duke University Press.

Kohli, Ajay (1989), "Determinants of Influence in Organizational Buying: A Contingency Approach," *Journal of Marketing*, 53 (July), 50-65.

Kotler, Philip (2003), *Marketing Management*, 11th Edition, Upper Saddle River, NJ: Prentice Hall.

Litman, Barry R. (1983), "Predicting Success of Theatrical Movies: An Empirical Study," *Journal of Popular Culture*, 16, 159-75.

Lynn, Michael and George M. Zinkhan and Judy Harris (1993), "Consumer Tipping: A Cross-Country Study," *Journal of Consumer Research*, 20 (3), 478-488.

Maddala, G.S. (1983), *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge, U.K.

May, Thomas (2004), "Political Authority in a Bio-terror Emergency," *Journal of Law, Medicine, and Ethics*, 32, 159-163.

Medved, Michael (1992), *Hollywood vs. America: Popular Culture & the War against Traditional Values*. New York: HarperCollins Publishers.

Myers, David G., and Helmut Lamm (1976), "The Group Polarization Phenomenon," *Psychological Bulletin*, 83, 602-627.

Morkes, John, Chen, Helen L., & Roberts, Donald F. (1997), "Adolescents' Responses to Movie, Television and Computer Game ratings and Advisories," *Paper presented at the annual meeting of the International Communication Association*, Montreal, Canada.

Neelamaghan, Ramya and Pradeep Chintagunta (1999), "A Bayesian Model to Forecast New Product Performance in Domestic and International Markets," *Marketing Science*, 18 (2), 115-136.

Olsberg/SPI and KEA (2003), "Empirical Study on Practice of the Ratings of Films Distributed in Cinemas Television DVD and Videocassettes in EU and EEA Member States," Prepared on Behalf of the European Commission, *Final Report*, 2003.

Peacock, Richard B. (2001), *The Art of Movie Making: Script to Scene*. Upper Saddle River, NJ: Prentice Hall.

Pechmann, Cornelia and Chuan-Fong Shih (1999), "Smoking Scenes in Movies and Antismoking Advertisements Before Movies: Effects on Youth," *Journal of Marketing*, 63 (July) 1-13.

Petty, Richard E. and John T. Cacioppo (1981), *Attitudes and Persuasion: Classic and Contemporary Approaches*. Dubuque, Iowa: Wm. C. Brown.

----and----(1986), *Communication and persuasion: Central and peripheral routes to attitude change*. New York: Springer-Verlag.

Powrie, Phil, Ann Davies, and Bruce Babington (2004), *The Trouble with Men: Masculinities in European and Hollywood Cinema*, Columbia University Press, New York.

Prentice, Robert. A. (2003), "Chicago Man, K-T Man, and the Future of Behavioral Law and Economics," *Vanderbilt Law Review*, 56 (6), 1661-1777.

Rantanen, Terhi (2005), *The Media and Globalization*, Sage Publications, London.

Ravid, S.Abraham (1999), "Information, Blockbusters, and Stars: A Study of the Film Industry," *Journal of Business*, 72 (4), 463-492.

Raz, Joseph (1986), *The Morality of Freedom*, Oxford University Press, New York.

Ridgeway, Cecilia L. (1983), *The Dynamics of Small Groups*. New York: St. Martin's Press.

Saks, Michael J. (1977), *Jury Verdicts*. Lexington, MA: D. C. Heath and Company.

Sawhney, Mohanbir S. and Jehoshua Eliashberg (1996), "A Parsimonious Model for Forecasting Gross Box-Office Revenues of Motion Pictures," *Marketing Science*, 15 (2), 113-131.

Schlesinger, Philip (1987), " On National Identity: Some Conceptualizations and Misconceptions Criticized", *Social Science Information*, 26 (2), 219-264.

Steenkamp, Jan-Benedict .E.M., Frenkel Ter Hofstede and Michel Wedel (1999), "A Cross-national Investigation into the Individual and National Cultural Antecedents of Consumer Innovativeness," *Journal of Marketing*, 63 (2), 55-69.

Stoner, James A.F. (1961), “A Comparison of Individual and Group Decisions Involving Risk,” Unpublished thesis, MIT, Cambridge, MA.

Sunstein, Cass, R. (2000), “Deliberative Trouble? Why Groups go To Extremes,” *The Yale Law Journal*, 110 (1), 71-120.

Swart, Christopher, William Ickes and Eric S. Morgenthaler (1978), “The Effect of Objective Self Awareness on Compliance in a Reactance Situation,” *Social Behavior and Personality*, 6 (1), 135 - 40.

Walsh, David and Douglas Gentile (2001), “A Validity Test of Movie, Television, and Video-game Ratings,” *Pediatrics*, 107 (6), 1302-1308.

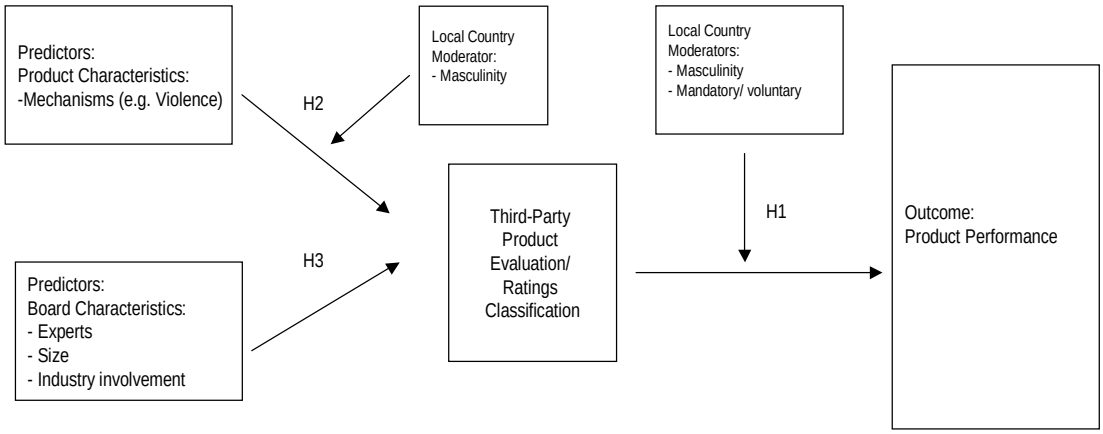

Figure 1: Conceptual Framework

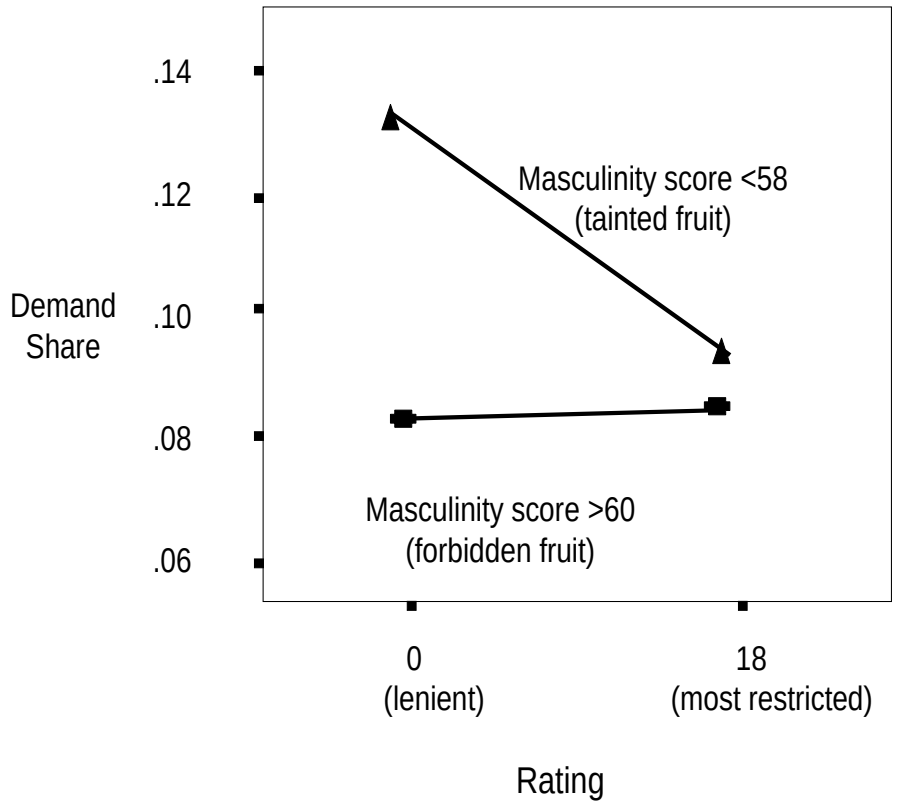

Figure 2: The Effects of the Ratings Classifications on Demand Share in Countries with Different Masculinity Scores

Note: There is a 'gap' in the masculinity scores around 60 in the sample (see Table 1). The masculinity score of 59 divides the countries into 5 more masculine and 3 less masculine (feminine) countries.

Table 1:
Descriptive Statistics of the Main Variables

	Aus	Fra	Ger	Hon	It	Spa	UK	US
Variable								
Mean Ratings Classification	12.99	1.55	10.76	10.38	2.69	7.70	12.81	14.14
Mean Demand Share	.08	.22	.08	.11	.07	.12	.09	.08
Mean Screen Share	.06	.14	.08	.11	.08	.09	.09	.08
Board Variables:								
Presence of Experts	No	Yes	No	No	Yes	No	No	No
Degree of Industry Involvement	0	3	2	0	0	0	1	1
Average Size of the Board	5	7	7	9	7	7	2	11
Country Variables:								
Masculinity Score	61	43	66	57	70	42	66	62
Mandatory Board Environment	Yes	Yes	No	Yes	Yes	Yes	No	No
Number of Movies	157	93	136	73	26	122	135	227

Range of Ratings Classifications	[0-18]
Range of Demand Share	[0- .84]
Range of Screen Share	[0- .32]
Range of Expert Involvement	[0-1]
Range of Industry Involvement	[0-3]
Range of Size of the Classification Board	[2-11]
Range of Masculinity	[42-70]
Mandatory System	[0-1]

Table 2: Correlations Matrix

	RAT	SSH	MAN	SIZ	EXP	IND	MAS
SSH	-.299						
	(.000)						
MAN	-.351	.055					
	(.000)	(.039)					
SIZ	.063	.065	-.112				
	(.051)	(.028)	(.000)				
EXP	-.548	.220	.450	-.031			
	(.000)	(.000)	(.000)	(.297)			
IND	-.260	.246	-.349	.029	.354		
	(.000)	(.000)	(.000)	(.331)	(.000)		
MAS	.408	-.227	.000	-.109	-.169	-.207	
	(.000)	(.000)	(.997)	(.000)	(.000)	.000	
DSH	-.249	.671	.098	-.022	.206	.228	-.228
	(.000)	(.000)	(.000)	(.459)	(.000)	.000	(.000)

RAT: Ratings Classification
SSH: Screen Share
MAN: Mandatory System
SIZ: Size of Classification Board
EXP: Expert Involvement
IND: Industry Involvement
MAS: Masculinity Index
DSH: Demand Share

Table 3:
The Effect of the Ratings Classifications on Demand Share:
Logistic Regression with Random Movie and Country Effects

Independent Variable	Hypothesis	Coefficient	t-value
Intercept		-4.4983	-77.36
Board Evaluation:			
Rating Classification (RAT)		-.0148***	-3.18
Strategic Marketing Variable:			
Screen Share (SSH)		23.9196***	37.41
Country Characteristics:			
Masculinity (MAS)		-.0093**	-2.32
Mandatory system (MAN)		.4801***	7.30
Interactions:			
RAT * MAS	H1a	.0014**	2.60
RAT * MAN	H1b	.0035	.30

Notes on Regression:
Dependent variable: Demand Share
-2LL = 2317. 9

Table 4:

Mean Rating Classifications for Identical Samples of Movies:

U.S. Vs Other Countries

Ratings Matched on Movies		F-value	Ratings Matched on Movies And Runtime		F-value
US	UK		US	UK	
13.91	12.81	F=4.17	14.32	13.31	F=2.02
(N=135)	(N=135)	(p=.042)	(N=62)	(N=62)	(p=. 158)
US	Aus		US	Aus	
13. 94	13.07	F=3.67	13.60	12.97	F=. 84
(N=156)	(N=156)	(p=.056)	(N=70)	(N=70)	(p.36)
US	Hon		US	Hon	
14.58	10.38	F=44.16	14.54	10.38	F=42.28
(N=73)	(N=73)	(p=.000)	(N=72[11])	(N=72)	(p=. 000)
US	Fra		US	Fra[12]	
14.04	1.55	F=466.24	14.13	1.82	F=365.49
(N=93)	(N=93)	(p=.000)	(N=79)	(N=79)	(p=. 000)
US	Ger		US	Ger	
13.87	10.76	F=31.14	13.73	10.41	F=28.66
(N=136)	(N=136)	(p=.000)	(N=103)	(N=103)	(p=. 000)
US	It		US	It	
14.54	2.69	93.65	14.39	3.02	F=12.16
(N=26)	(N=26)	(p=.000)	(N=23)	(N=23)	(p=. 003)
US	Spa		US	Spa[13]	
13.87	7.70	F=74.74	13.76	7.36	F=13.68
(N=122)	(N=122)	(p=.000)	(N=25)	(N=25)	(p=001)

[11] Only the movie *Face/ Off* had a different runtime in Hong Kong compared to the U.S.

[12] We checked for possible effects of language dubbing, which may not be visible in the runtime. We deleted all movies that have 'language' as a mechanism that may have been dubbed in the foreign country. This additional restriction on the sample did not result in substantial differences in results for any comparison.

[13] In Spain, the length of the movie was given in meters and we had to estimate the runtime.

Table 5:
Drivers of the Ratings Classification (Dependent Variable):
Linear Regression with Random Movie and Country Effects and Ordered Probit

Independent Variable	Hyp	Linear Regression with Random Effects Coef.	t-value	Ordered Probit Coef.	z-value
Constant		-3.2723***	-6.97		
Movie's Mechanisms:					
Violence	H2a	3.1135***	10.74	0.6840***	5.28
Sex		1.5661***	4.54	0.3239***	2.86
Drugs		1.0337**	1.74	0.3645*	1.56
Nudity		.7402*	1.62	0.2890*	1.45
Rough humor		-.6538	-1.07	-0.2699	-1.17
Language		1.1888***	4.02	0.2632**	1.96
Gore		2.9182***	4.97	0.5849***	2.33
Thematic elements		- .3259	-.75	-0.1777	-0.91
Sensuality		-1.7287***	-3.42	-0.6234***	-3.99
National Culture Characteristics:					
Masculinity		.1743***	10.10	0.01096***	6.65
Violence * Masculinity	H2b	-.0390*	-1.28	-0.00633***	-2.66
Board Characteristics:					
Experts	H3a	-7.8592***	-14.00	-2.34104***	-10.21
Industry involvement	H3b	-.0797	-.45	0.00841	0.19
Board size	H3c	.1929***	3.95	0.01339**	1.65

Notes on Mixed Linear Regression:
Dependent variable: recoded rating var.

-2RLL = 5534.9

Notes on Ordered Probit:
Dependent variable: un-recoded rating var.
Parameters are Robust estimates using Cluster (Movie) in Stata7

χ^2 (14) = 314.54***
Log Likelihood = -2027.39
Pseudo R^2= .12

*p<. 1 **p<. 05 ***p<. 01 (one-tailed)

Appendix 1: Local Ratings in the Sample and their Corresponding Numerical Scales

Country	Local Rating:	Numerical Scale Used:
US	G	0
	PG	10[14]
	PG13	13
	R	17
Aus	G	0
	PG	10
	M/MA	15
	R	18
Fra	U	0
	-12	12[15]
Ger	U	0
	6	6
	12	12
	16	16
	18	18
Hon	I	0
	IIa	10
	IIb	13
	III	18
It	T	0
	VM14	14
Spa	T	0
	7	7
	13	13
	18	18
UK	U	0
	PG	10
	12	12
	15	15
	18	18

[14] PG ratings do typically not specify a cut-off age. However, telephone interviews suggest that the local boards employ the age of 10 as a practical benchmark (AUS, UK, US).

[15] –16 (and –18) exist but are not in the sample. Less than 5% of *all* movies submitted to the classification organization obtained a –16 or –18 rating in France in 2003.