Consensus and Contrasts in Consumers’ Cinematic Assessments:
Gender, Age, and Nationality in Rating the Top-250 Films

Dean Keith Simonton
University of California, Davis

Jeremy J. Graham and James C. Kaufman
Learning Research Institute, California State University, San Bernardino

Motion pictures provide among the most conspicuous manifestations of worldwide popular culture. One specific manifestation of this universal presence appears in the cinematic assessments compiled and updated on Internet Web sites. This empirical inquiry investigated the consumer ratings that the Internet Movie Database used to determine the “Top-250” all-time great movies. Of particular interest was how these ratings depended on gender (male vs. female), age (under 18, 18–29, 30–44, and 45 or over), and nationality (U.S. vs. non-U.S. voters). In addition, the investigation explored how any evaluation discrepancies in these three demographic categories might be attributed to year of release (e.g., classic vs. contemporary films), movie honors (viz., Oscar vs. non-Oscar nominations and awards), and the MPAA rating (R, PG-13, PG, and G). Correlational, principal components, and multiple regression analyses indicate the following core conclusions. First, a very broad and impressive consensus permeates all evaluations no matter what the gender, age, or nationality contrasts. Second, although gender and nationality both exhibit contrasting assessments, age provides the main contrast that supports departures from the consensus: Those under 30 have strikingly different assessments than those 30 and over. Third and last, although movie awards and MPAA ratings clearly have a role to play in these differences, the year of release was by far the most critical predictor. Older consumers prefer older movies while younger consumers prefer movies that are more recent. After some conjectures regarding the reasons for this pronounced contrast, the discussion closes by mentioning the dynamic nature of these popular ratings.

Keywords: film ratings, consumers, gender, age, nationality

Mainstream movies certainly represent among the most prominent and enduring expressions of popular media culture (Simonton, 2011). Beginning as a minor form of entertainment in the late 1890s and early 20th century—as the once ubiquitous Nickelodeons well exemplify—film quickly expanded into a major high-visibility event as well. This expansion accelerated once Hollywood emerged as a conspicuous world filmmaking center—especially after its famed studios developed the “star system” that created the first universally recognized and idolized celebrities. Fans flocked to the local movie theater to see their favorite male and female actors perform in their latest motion picture. Although the film industry and its stars lost some ground to other media, especially TV, pop music, video games, and the Internet, the advent of a blockbuster movie continues to grab global attention. According to the Internet Movie Database, the 2009 movie Avatar opened on 3,452 screens, took in over $77 million on the opening weekend, and grossed almost $2.8 billion worldwide. It probably would be difficult to find anybody in the developed world who has never heard of the movie—even if that person refused to see it. Films like Avatar become embedded in popular culture.
Cinematic prominence is apparent in other media beyond the theatrical releases. To offer the most obvious examples: (a) movies provide common viewing fare on network, cable, and satellite TV programming; (b) they are distributed as DVDs through video stores, the mail, the Internet, and the local supermarket; (c) they can be streamed or downloaded (legally or illegally) to be watched on a computer; (d) they support a large entourage of film critics who publish their cinematic assessments in newspapers, magazines, radio, TV, and Internet media; (e) they inspire abundant print and electronic media that cover the scandalous life stories and “red carpet” fashion trends of industry celebrities; (f) their plots and characters are sometimes closely linked to the comics, graphic novels, video games, and bestselling books; and (g) they motivate the creation of innumerable Web sites devoted to films and the talents involved in filmmaking. The last manifestation includes Web sites devoted to the assessment of cinematic merit, including sites that allow cinema consumers to report their own judgments on the best and worst in film. These consumer ratings provide an alternative perspective to the judgments of film critics (e.g., www.metacritic.com), industry professionals (e.g., www.oscars.org), and movie journalists (e.g., www.goldenglobes.org)—none of which have to correlate highly with consumers’ cinematic assessments (Plucker, Holden, & Neustadter, 2008; Plucker, Kaufman, Temple, & Qian, 2009). Consumers and critics, in particular, do not necessarily apply the same criteria when judging films (Holbrook, 1999). For complex reasons, too, consumer cinematic assessments need not even have strong associations with box office performance (Simonton, 2011; e.g., Plucker et al., 2008). Accordingly, such ratings introduce an independent take on cinematic impact.

Perhaps the most popular Internet consumer assessments are the “user ratings” compiled at the Internet Movie Database (IMDb.com). Here “regular voters” can assess films on a 10-point scale. Thousands upon thousands of films have undergone such evaluations, and the number of raters can become truly prodigious. For example, the 2008 The Dark Knight has been rated by more than a half-million consumers. Besides providing an overall evaluation, the Web site offers breakdowns by gender, age, and nationality (viz., U.S. vs. non-U.S. voters). These breakdowns provide an opportunity to investigate how consumer tastes vary across different demographics. Although previous research has shown that critic evaluations exhibit an exceptional consensus on the relative merits of various films (Boor, 1990; Plucker et al., 2008; Simonton, 2007), this consensus is misleading to the extent that film critics are predominantly older males operating for mainstream media in the United States.

Hence, our goal in this exploratory investigation is to assess gender, age, and nationality contrasts in consumers’ cinematic assessments. Do men and women rate films differently? Do teenagers offer different judgments than adults? Do U.S. consumers offer cinematic opinions that differ from non-U.S. consumers? If the answer to any one of these questions is affirmative, then the follow-up question becomes, what are some of the foundations for the differences? In this exploratory inquiry, we will examine three potential sources for assessment contrasts, namely, (a) the date of theatrical release date, (b) the reception of major movie nominations and awards (Oscars and otherwise), and (c) the mature-content ratings put out by the Motion Picture Academy of America (MPAA). For instance, do younger consumers rate older film classics less highly than do more mature consumers? Are there gender differences in preferences for films containing lots of “sex and violence” as indicated by an R(restricted) MPAA rating? Finally, are U.S. consumers inclined to give more weight for the Oscars bestowed by the American Academy of Motion Picture Arts and Sciences? These three sources are easily assessed for large numbers of films and at the same time have been shown to have predictive value for other criteria of cinematic impact, including box office performance and critical acclaim (Simonton, 2009b).

To keep this investigation manageable, we decided to restrict the sample of films to those that IMDb.com identifies as the “Top-250” movies according to consumer assessments. Because these films tend to be very well known (at least to the cinephiles who are more likely to frequent IMDB), they are most likely to have been viewed and rated by the largest number of moviegoers—and thus, present a more reliable and valid indicator of consumer tastes than if we included more esoteric fare. To be sure, because we will be thus truncating the sample to the upper end of the popularity distribution, the
variance in the assessments will be reduced. Yet this variance reduction will bias our study from obtaining any significant results because the effect sizes will be attenuated. If we can identify influential factors for the Top-250 movies, then it should be even easier to do so for films that range from the masterpieces to the turkeys (Simonton, 2007).

**Method**

The sample consisted of the 250 films identified as the highest rated as of September 20, 2011. According to IMDb.com, a "true Bayesian estimate" is calculated using the following formula:

\[ WR = \left( v \div (v + m) \right) \times R + \left( m \div (v + m) \right) \times C, \]

where \( WR \) is the weighted rating, \( v \) the number of votes the movie received, \( m \) the minimum number of votes required to be considered for inclusion in the Top 250 (\( m = 3,000 \)), \( R \) the average (or mean) rating that the movie received, and \( C \) "the mean vote across the whole report (currently 6.9)." Only the assessments of "regular voters" are included in the calculations (a restriction that reduces the opportunities for focused campaigns to promote or demote a film's placement on the list). Throughout this article, we will refer to \( WR \) as the overall cinematic assessment. In addition, IMDb.com breaks the evaluations by (a) gender (Male/Female), (b) age (under 18, 18–29, 30–44, and 45 and older), and nationality (U.S./Non-U.S.). These will later be used to generate a set of contrast measures. In addition, for one later analysis we will have occasion to use assessments broken down by both gender and age, yielding eight assessments.

IMDb.com also provided the following information: (a) release year (the time of the domestic theatrical release), (b) Oscar awards and nominations as well as non-Oscar awards and nominations, and (c) the MPAA rating of the domestic theatrical release. Release year was taken unchanged, a variable that ranged from 1924 (Sherlock Jr., a Buster Keaton film) to 2011 (Harry Potter and the Deathly Hallows: Part 2), with \( M = 1,978.40 \) and \( SD = 24.25 \). The breakdown by decade was as follows: 1920s, 2.8%; 1930s, 4%; 1940s, 7.3%; 1950s, 13.6%; 1960s and 1970s, 9.6% each; 1980s, 11.2%; 1990s, 15.2%; 2000s, 23.6%; and 2010s, 3.2%. Hence, although more recent films are strongly represented, older films are not completely overlooked, including classic films from the silent era.

Unlike release year, the other measures had to undergo additional data transformations.

In the case of the awards and nominations, these were summed to produce two indicators of movie honors: Oscar honors (\( M = 5.67 \), \( SD = 5.84 \), range = 0–24) and Non-Oscar honors (\( M = 56.30 \), \( SD = 58.05 \), range = 0–302). In effect, the resulting indicators give awards twice the weight of nominations (Simonton, 2004a, 2009a). This procedure is common practice in previous research in which awards were used to predict either box office performance or film critic evaluations (e.g., Basey, Chatterjee, & Ravid, 2003; Ginsburgh, 2003; Sochay, 1994). Because Harry Potter and the Deathly Hallows: Part 2 had not yet undergone the nomination/award process by the time this study was conducted, \( n = 249 \) for analyses involving these variables.

Lastly, the MPAA ratings were coded into the following four 0–1 dummy variables: \( R \) (restricted), \( PG-13 \) (parental guidance, some material not appropriate for viewers younger than 13), \( PG \) (parental guidance), and \( G \) (general admission). These ratings are not available for 32% of the films—either film classics or foreign movies. The percentages receiving ratings were as follows: 40.8% \( R \), 10.0% \( PG-13 \), 12.4% \( PG \), and 4.4% \( G \). As often the case in top-flight movies, R-rated films dominate the sample (e.g., Simonton, 2005b; see also De Vany & Walls, 2002).

**Results**

The data analyses fall naturally into four groups: basic statistics, principal components analysis, correlation coefficients, and multiple regression.

**Basic Statistics**

Table 1 presents the basic statistics for the overall assessments and the assessments broken down by gender, age, and nationality. The means range from 8.055 to 8.435, a relatively small degree of variation. However, the standard deviations range from 0.246 to 0.617, suggesting considerable differences in the mag-
The ratings by voters under 18 were especially variable. This conclusion is reinforced by the ranges. Although the overall assessments have a minimum score of 7.9, subgroups dip down much lower, namely, 6.5 for those 45 and over, 6.4 for the females, and 5.5 for those under 18. For the last age group, some of the films in the Top 250 are considered barely above mediocrity. Also noteworthy is the fact that the maximum scores range from 9.1 to 9.4, the higher scores being favored by the younger voters. Hence, no movie gets a perfect 10 from all raters. On this score, consumers contrast with film critics who will occasionally concur that a particular movie deserves a perfect rating (Simonton, 2002). An example is the 1975 One Flew Over the Cuckoo’s Nest that earned “five stars” or the equivalent in all popular movie guides (Simonton, 2004b), as well as sweeping the Oscars in the five primary categories (Picture, Director, Screenplay, Actor and Actress). Yet as noted earlier, critics form a more homogeneous demographic group than do consumers.

It should be reported that (a) the male and female assessments correlate .400 \((p < .001)\); (b) the four age category assessments correlate between .058 (under 18 with 45 and over; \(p = .364\)) and .780 (18–29 with 30–45; \(p < .001\)); and (c) the U.S. and non-U.S. assessments correlate .576 \((p < .001)\). Hence, the consensus in each case is far from strong enough to rule out contrasts (see also Simonton, 2009c). The disagreements between the youngest and oldest raters are the most salient—a point that will reitered in later sections.

Table 1 also gives the correlations between the quantitative rating and the rank based on that rating. Because of the loss of information in converting from an interval to ordinal scale, the correlation cannot be perfect, as the \(-.932\) shows (and, of course, ranks are inversely related to ratings). More significant is the fact that these correlations vary greatly across gender, age, and nationality. This variation partially reflects the ranges within each category. For example, if females provide more variable ratings than males, then their ratings might correspond less closely to the overall assessment. Nonetheless, it is also true that these discrepancies can be ascribed to differences in participation, males voting more than females. Table 1 provides the proportion of voting participants who fall in each demographic category. It is apparent that the higher the percentage, the stronger is the association with the overall ranking. More precisely, the correlation between these two columns is an impressive \(-.839\) \((n = 9, p < .01)\). Thus, because males outnumber the females, the young outnumber the mature, and non-U.S. voters outnumber U.S. voters, the males, young, and non-U.S. voters have the biggest impact on the ranking.

Not recorded in Table 1 is the correlation of .631 \((p < .001)\) between the overall assessment and the total number of voters. The higher rated movies attract the most votes. A similar corre-
lation holds for film critics; critically acclaimed movies receive the most film reviews \( (r = .358, N = 1,514, p < .001; \) using Metacritic.com data collected for studies reported in Simonton, 2009a, 2009c). Together these correlations illustrate a bandwagon effect for both consumers and critics, but with the effect far stronger for the former.

**Principal Components Analysis**

To get a better idea of the nature of the ratings, we conducted a principal components analysis of the assessments broken down by gender and age. Because nationality was necessarily omitted, these eight judgments are completely independent, the rating of any given user appearing in one and only one category. Significantly, just three components had eigenvalues exceeding one, and the loadings were substantively interpretable without the need to impose a rotation. The results appear in Table 2.

The first component accounts for more than 46% of the total variance, and features a high positive loading from all eight assessments—albeit the highest loadings are for males 18–29 and females 30–44. This component clearly represents an overall cinematic greatness factor that compares favorably with the single-factor consensus commonly witnessed in other multiple-indicators of popularity or eminence (Simonton, 1991, 1998). For example, the consumer consensus is comparable to that seen in the evaluations of film critics and the honors bestowed by professional organizations on the films released each year (Simonton, 2004a, 2004b).

The remaining two components represent contrasts rather than consensus. The second component pits younger raters against the mature raters, with age 30 providing the pivotal point with raters under 18 strongly disagreeing with raters 45 and older. This division also falls in line with the percentages given in Table 1, where 51% are under 30 and 49% are 30 or over. Even so, this component accounts for less than half the variance explained by the first component.

The third component explains even less, around 15%, and is clearly distinct from the previous two. The female assessments all have positive loadings and the male assessments negative or near zero loadings. Although less well differentiated, this component obviously embodies a male-female contrast factor, with an emphasis on how the women differ from the baseline established by the men.

Notably, the two contrast components together explain appreciably less variance in the assessments than does the first consensus component \( (i.e., 46.3–21.5–15.3 = 9.5). \) Consensus dominates over contrasts. The same has been shown to hold for the judgments of film critics (Simonton, 2009c).

**Correlation Coefficients**

Now that we know the pivoting point for the age assessments (viz., age 30), we generated three contrast variables: (a) *Male versus Female* (MvF; male minus female assessments), (b) *Young versus Mature* (YvM; assessments for ages 29 and below minus assessments for ages 30 and above), and (c) *U.S. versus Non-U.S.* (USvNonUS; assessments of U.S. voters minus the assessments of voters outside the U.S.). These three new variables, along with the overall assessment, could then be correlated with release year, Oscar and Non-Oscar honors, and the four MPAA dummy variables: R, PG-13, PG, and G. The resulting correlations are given in Table 3.

The differences are quite pronounced. The overall assessment has only two correlates: the two nomination/award indicators. Honored movies do tend to be rated more highly by IMDb.com voters. Yet the latter do not have any special bias with respect to release year and MPAA ratings. In comparison, we can note the following about the three contrast variables.

### Table 2

**Principal Components Analysis: Cinematic Assessments by Gender and Age Groups**

<table>
<thead>
<tr>
<th>Assessments</th>
<th>I</th>
<th>II</th>
<th>III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males under 18</td>
<td>.733</td>
<td>-.460</td>
<td>.015</td>
</tr>
<tr>
<td>Females under 18</td>
<td>.563</td>
<td>-.400</td>
<td>.552</td>
</tr>
<tr>
<td>Males 18–29</td>
<td>.802</td>
<td>-.294</td>
<td>-.459</td>
</tr>
<tr>
<td>Females 18–29</td>
<td>.774</td>
<td>-.225</td>
<td>.300</td>
</tr>
<tr>
<td>Males 30–44</td>
<td>.737</td>
<td>.130</td>
<td>-.616</td>
</tr>
<tr>
<td>Females 30–44</td>
<td>.802</td>
<td>.224</td>
<td>.174</td>
</tr>
<tr>
<td>Males 45 and over</td>
<td>.482</td>
<td>.806</td>
<td>-.130</td>
</tr>
<tr>
<td>Females 45 and over</td>
<td>.438</td>
<td>.704</td>
<td>.435</td>
</tr>
</tbody>
</table>

% total variance explained: 46.3, 21.5, 15.3

*Note. N = 250.*
First, gender contrasts are quite prominent. Females, relative to males, rate more highly older releases (film classics), films that receive movie honors, Oscar or otherwise, and movies that receive MPAA ratings of either PG-13 or G.

Second, age contrasts are also conspicuous, but following a different pattern from the gender contrasts (as would be expected from the orthogonal nature of components two and three). The young prefer movies that are more recent, movies that receive honors besides the Oscars, and movies that receive the two most mature MPAA ratings, R and PG-13. Only the more mature raters appear to give any credence to the Oscar awards and nominations. The last finding is compatible with the high correspondence between Oscar honors and the evaluations of film critics (Simonton, 2004a).

Third, nationality does play a role as well, at least as defined as the U.S. versus non-U.S. contrast. Outside the U.S., moviegoers prefer older and R-rated films, whereas inside the U.S. they prefer films that receive Oscar awards or nods and that are rated PG (cf. Cerridwen & Simonton, 2009).

Before we make too much of the above findings, we must acknowledge that some of these correlation coefficients might be spurious because some variables share appreciable variance. Especially notable are historical trends introducing correlations with release year. Specifically, the latter variable correlates positively with Non-Oscar honors ($r = .585, p < .001$) and MPAA ratings of R ($r = .520, p < .001$) and PG-13 ($r = .253, p < .001$). In other words, the number of potential nominations and awards beyond the Oscars has tended to increase over the years (owing to an increase in the number of award ceremonies), and the film content has become increasingly aimed at mature audiences (see also Simonton, Skidmore, & Kaufman, in press, who found similar increases for family films).

### Multiple Regression

To adjust for shared variance among the variables, the four cinematic assessments—overall and the three contrasts—were regressed on the correlates in Table 3. The standardized partial regression coefficients are found in Table 4.

Although several variables no longer feature significant effects, significant predictors still emerged, some with even bigger effects (owing to statistical suppression; Maassen & Bakker, 2001). First, in the case of overall, male versus female, and U.S. versus Non-U.S. assessments the relations are negative, whereas for the young-versus-mature contrast the relation is positive. That is, older movies are rated more highly than recent movies, particularly by females, the mature, and those living outside the United States. Stated differently, young U.S. males like more recent motion pictures. Film honors display a contrasting pattern. Although U.S. consumers are impressed with Oscar non-

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Overall</th>
<th>MvF</th>
<th>YvM</th>
<th>USvNonUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Release year</td>
<td>$-0.226^*$</td>
<td>$-0.227^*$</td>
<td>$0.670^{***}$</td>
<td>$-0.321^{**}$</td>
</tr>
<tr>
<td>Oscar honors</td>
<td>$0.140$</td>
<td>$-0.153$</td>
<td>$-0.128^*$</td>
<td>$0.290^{***}$</td>
</tr>
<tr>
<td>Non-Oscar honors</td>
<td>$0.193^*$</td>
<td>$0.007$</td>
<td>$-0.039$</td>
<td>$0.130$</td>
</tr>
<tr>
<td>MPAA: R</td>
<td>$0.260^*$</td>
<td>$0.214^*$</td>
<td>$0.046$</td>
<td>$0.032$</td>
</tr>
<tr>
<td>MPAA: PG-13</td>
<td>$0.137$</td>
<td>$-0.003$</td>
<td>$0.039$</td>
<td>$0.046$</td>
</tr>
<tr>
<td>MPAA: PG</td>
<td>$0.091$</td>
<td>$0.036$</td>
<td>$0.054$</td>
<td>$0.163^*$</td>
</tr>
<tr>
<td>MPAA: G</td>
<td>$0.057$</td>
<td>$-0.095$</td>
<td>$0.017$</td>
<td>$0.081$</td>
</tr>
<tr>
<td>$R^2$</td>
<td>$0.094^{**}$</td>
<td>$0.088^{**}$</td>
<td>$0.487^{***}$</td>
<td>$0.209^{***}$</td>
</tr>
</tbody>
</table>

**Note.** Because one 2011 film had not yet completed the award season at the time of the investigation, $n = 249$.  
* $p < .05$.  
** $p < .01$.  
*** $p < .001$. 

### Table 3

**Pearson Product-Moment Correlation Coefficients**

<table>
<thead>
<tr>
<th>Correlates</th>
<th>Cinematic assessments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overall</td>
</tr>
<tr>
<td>Release year</td>
<td>$0.061$</td>
</tr>
<tr>
<td>Oscar honors</td>
<td>$0.232^{***}$</td>
</tr>
<tr>
<td>Non-Oscar honors</td>
<td>$0.224^{***}$</td>
</tr>
<tr>
<td>MPAA: R</td>
<td>$0.096$</td>
</tr>
<tr>
<td>MPAA: PG-13</td>
<td>$0.067$</td>
</tr>
<tr>
<td>MPAA: PG</td>
<td>$-0.005$</td>
</tr>
<tr>
<td>MPAA: G</td>
<td>$-0.026$</td>
</tr>
</tbody>
</table>

**Note.** MvF = male minus female assessments, YvM = under than 29 assessments minus older than 29 assessments, and USvNonUS = US assessments minus non-US assessments. $N = 250$ except for the award/nomination correlations because one 2011 film had not yet completed the award season at the time of the investigation.  
* $p < .05$.  
** $p < .01$.  
*** $p < .001$. 

---

Because one 2011 film had not yet completed the award season at the time of the investigation, $n = 249$.  
* $p < .05$.  
** $p < .01$.  
*** $p < .001$. 

---

Because one 2011 film had not yet completed the award season at the time of the investigation, $n = 249$.  
* $p < .05$.  
** $p < .01$.  
*** $p < .001$. 

---

Because one 2011 film had not yet completed the award season at the time of the investigation, $n = 249$.  
* $p < .05$.  
** $p < .01$.  
*** $p < .001$. 

---

Because one 2011 film had not yet completed the award season at the time of the investigation, $n = 249$.  
* $p < .05$.  
** $p < .01$.  
*** $p < .001$. 

---

Because one 2011 film had not yet completed the award season at the time of the investigation, $n = 249$.  
* $p < .05$.  
** $p < .01$.  
*** $p < .001$. 

---

Because one 2011 film had not yet completed the award season at the time of the investigation, $n = 249$.  
* $p < .05$.  
** $p < .01$.  
*** $p < .001$. 

---

Because one 2011 film had not yet completed the award season at the time of the investigation, $n = 249$.  
* $p < .05$.  
** $p < .01$.  
*** $p < .001$. 

---

Because one 2011 film had not yet completed the award season at the time of the investigation, $n = 249$.  
* $p < .05$.  
** $p < .01$.  
*** $p < .001$.
inoninations and awards, young consumers are not. In contrast, the Non-Oscar honors are only associated with the overall assessments and thus, are not differentiated according to gender, age, or nationality. Finally, regarding the MPAA ratings we find that R-rated films receive higher overall assessments as well as higher assessments from males relative to females. U.S. consumers also like PG films more than non-U.S. consumers do.

The final row of Table 4 gives the proportion of assessment variance accounted for by the predictors. Almost 10% of the overall assessment is explained by the predictors, a figure that is slightly reduced for the male versus female contrast. Yet the proportion of variance more than doubles for the U.S. versus Non-U.S. contrast and multiples almost five times for the young versus mature contrast. Hence, in these data, the biggest contrasts involve age, with nationality a distant second. Indeed, it must be deemed striking that almost half of the variance in the young versus mature difference can be so explained. More striking, still, is the fact that young consumers tend to dislike old movies, just as they tend to dislike movies featured on Oscar night.

Discussion

It cannot be overemphasized that this inquiry was exploratory only. The aim was merely to explore the nature of consumers’ cinematic assessments, as represented by the IMDb.com user ratings for the Top 250 films. Besides scrutinizing the overall evaluations, we were interested in gender, age, and nationality contrasts in film preferences. Moreover, we have managed to pin down some of the reasons for the contrasts, at least insofar as the differences reflect contrary responses to release date, movie awards, and MPAA ratings. Perhaps the most fascinating findings are three in number.

First, notwithstanding the contrasts, the consensus regarding consumers’ cinematic assessments is actually quite large, largely replicating what was found for film critics as well (Plucker et al., 2008; Simonton, 2009c). As seen in Table 2, the first principal component accounts for 46% of the total variance in assessments broken down by age and gender. The remaining young versus mature and male versus female components taken together explain much less variance than this first component alone.

Second, when we investigate the three contrasts—gender, age, and nationality—age emerges as the most critical. Again, in Table 2, the 29 and younger raters are diametrically opposed to the 30 and older raters. Furthermore, as reported earlier, the ratings by voters under 18 correlate only .058 with voters who are 45 and over. Curiously, judging by the loadings in Table 2, this age differential in evaluations is accentuated somewhat by gender. Males under 18 have the most negative loading and males 45 and over have the most positive loading.

Third, the movie attribute that most strongly supports the observed contrasts is the film’s year of theatrical release. This predominance is most salient in Table 4, which provides the standardized partial regression coefficients. Not only is release year the only predictor that predicts the overall assessment and all three contrasts, but also release year has the highest predictive value in all but one equation (viz., for overall). Release year has the strongest predictive power in the case of the age contrast, a contribution that enables the equation as a whole to account for 49% of the variance. The biggest take-home message here is that young voters like recent films while mature voters like old films. But why?

Without additional data, we can only speculate. One possibility is that this represents a mere age effect. Older films require more maturity to appreciate because they feature superior acting, writing, and direction, the primary qualities of great films (Simonton, 2011). Movies that are more current, in contrast, may substitute special effects for substance, and all the film’s assets are worn on the surface. If so, the younger voters might alter their cinematic assessments with increased age, finally agreeing that the classics are indeed classics once they reach middle age. Such an argument would also explain the negative relationship between a film winning an Oscar and appealing to younger voters.

Another reasonable explanation is that the age-date correlation reflects a cohort effect. Perhaps the MTV generation perceives the more art-oriented films of the past as pretentious and boring “old fogy” stuff. In that case, even after these younger voters grow up to wear reading
glasses, they will still prefer more recent faire—movies that seem more similar to other youthful media like video games rather than outdated forms like live theater.

Yet another potential cause would not be age per se but rather whether or not the voter has some claim to being a film professional rather than just an amateur viewer. Although not all mature voters would have serious involvement with the film industry and its spin-offs, certainly the proportion of such voters would increase from the youngest to the oldest age categories. Only child performers would be high school age or younger. One asset of this third explanation is that it fits the tendency for the older voters to like films that receive Oscar nods and statuettes. After all, the Academy Award decisions are made by approximately 6,000 of the top professionals in the movie business. Even if a voting member of the Academy were not a “regular voter,” many IMDb regular voters would certainly have a great deal of respect and admiration for members of the Academy.

Unfortunately, we cannot really address this issue without collecting considerably more data, and particularly data collected longitudinally to permit studying changes in preferences over time. In addition, it would be very useful to add more variables, especially box office and critical evaluations. We could not incorporate these variables in this study because a very large proportion of the 250 appeared before such data were readily available (see, e.g., Simonton, 2005a). Nevertheless, if the sample size is increased while deleting older films, it should be easy to examine the same cinematic assessments using a larger number of potential predictors. Such comprehensive inquiries would allow us to do for these consumer assessments what has already been accomplished for financial performance, critical acclaim, and film honors, the three major criteria of cinematic impact (Hadida, 2008; Simonton, 2009b).

The three cinematic impact criteria just mentioned suggest a fundamental discrepancy with the consumer assessments. Box office, critic ratings, and movie awards eventually settle down to a stable, even permanent judgment of a film’s impact. After a film completes its theatrical release, after the critics have moved on to rating more recent films, and after all of the award ceremonies for its release year have run their course, it obtains its final scores on all three criteria. Naturally, some of the classic films—such as Walt Disney’s old animated features or the Star Wars films—might be re-released for the consumption of more recent generations. Also, from time to time some entrepreneurs will put out a “director’s cut” or “special edition” on DVD or Blu Ray. Even so, the already established figures for box office, critical evaluations, and movie awards will remain unaltered and static within a year’s time of its opening weekend.

In contrast, IMDb never closes the polling booth on its Top-250 voters. As a result, the ranking is dynamic with respect not just to a film’s ordinal placement, but also regarding to whether the film continues on the list. Indeed, the Top-250 when we began collecting the data for this investigation has already changed. A portion of this change can be attributed to the simple addition of recent releases. In the sample for our investigation, only one 2011 film appeared on the list, Harry Potter and the Deathly Hallows: Part 2, which stood at rank #111. But less than a handful of months later, two additional films appeared: Drive at rank #137 and A Separation at rank #174, with the last Harry Potter installment bumped down between the two at rank #160. It is understood that if two movies are added then two movies must be dropped, and given the predilections of the main voters, that implies older films will be knocked off at the bottom. Confirming expectation, we identified the 1937 Grand Illusion (formerly #249) and the 1944 Arsenic and Old Lace (formerly #250) as the two unlucky films to go into Top-250 oblivion—probably never to return owing to the constant influx of films more flashy and contemporary.

Admittedly, a strong constraint is imposed on this revision process so that the list cannot be easily turned topsy-turvy: The highest ranked films have their position predicated on so many votes—hundreds of thousands—that it would be most difficult to revamp the ranking from scratch. Even small revisions would be very difficult. To illustrate, consider the film in the top spot: the 1994 Shawshank Redemption. Although this film is certainly very good, perhaps even great, should it really be #1? The regression results reported in Table 4 suggest not. When the overall assessment is regressed on all of the predictors, one and only one film appears as an outlier, with the large Studentized residual
of 3.761—the very same movie. In concrete terms, Shawshank Redemption has a raw residual of 0.916, indicating that it is rated almost one point too high on a 10-point scale, an error of almost 10%. Rather than being placed just above the 1972 The Godfather and the 1974 The Godfather Part 2, it should be placed roughly at the level of the 1980 Raging Bull or the 1992 Unforgiven. The oddly high rank of Shawshank Redemption has been noted in the popular press; one blog considers the choice one of 250 reasons not to trust IMDb (Ziggi, 2011) and another blogger calls the #1 pick one of the great mysteries of our time (Sexton, 2008).

Judging from the equation, Shawshank Redemption is overrated because it earned fewer Oscar and non-Oscar honors than might be expected for a film so highly praised by IMDb’s regular voters. Even so, nothing can really be done about this evaluative inaccuracy because Shawshank Redemption received more votes than any other movie in the Top-250—fully 664,154 at the time of writing this sentence. That gives it far too much inertia to budge it from its exaggerated status on the list. It is literally stuck at the top by popular acclaim. All that said, it remains true that consumer ratings like the IMDb Top-250 have a finger on the pulse of popular media culture. As such, future research needs to understand in more detail the psychology of the voters and the process that discriminates the best from the very best in a particular moment of cinematic history. This inquiry is only a first start.

References


Received November 23, 2011
Revision received November 23, 2011
Accepted January 6, 2012