Visual Distortion Bias in Consumer Choices

Tao Lu, Alex Wang, May Yuan, Xiaoquan (Michael) Zhang

Abstract

Existing research on word-of-mouth considers various descriptive statistics of rating distributions, such as the mean, variance, skewness, kurtosis, and even entropy and the Herfindahl-Hirschman Index. But real-world consumer decisions are often derived from visual perceptions about displayed rating distributions in the form of histograms. In this study, we argue that such distribution charts may inadvertently lead to a consumer-choice bias that we call the Visual Distortion Bias (VDB). We propose that consumers have a tendency to be misled by salient features of distributions in visual decision-making. In an illustrative model, we derive a measure of the VDB. In a series of experiments, we identify the VDB's significant impact on consumer choices. We show that with the VDB, consumers may make choices that violate widely accepted decision rules. In our experiments, subjects are observed to prefer products with lower average ratings. They violate widely accepted modeling assumptions, such as branch independence and first-order stochastic dominance.

Keywords: Online Ratings, Online WOM, Graphical Decision Support, Visual Bias, Decision under Uncertainty

Forthcoming: Management Science

1 Introduction

Consumer decision-making is often made easier by visual aids, such as histograms or scatter plots. Many online rating platforms use rating distribution charts to offer more information than simple summary statistics such as the mean and the variance (Moore and Lafreniere, 2020). Internet Movie Database (IMDB), for example, provides a distribution chart, in addition to rating volume, valence, topics, and ranked text reviews, in its consumer-review section (see Figure 1).[1]

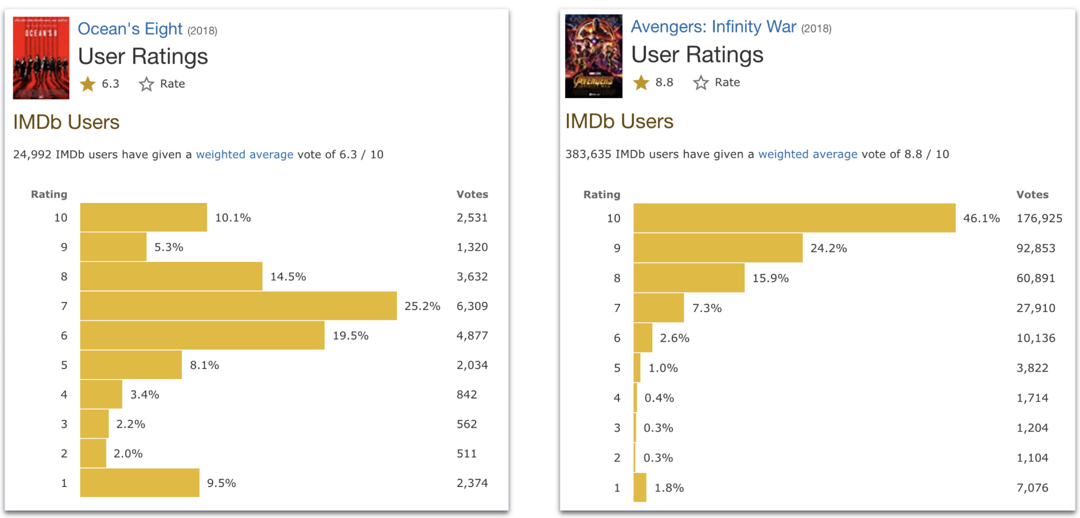

Figure 1: IMDB's Rating Distribution Charts

Despite the broad use of distribution charts in online rating systems and their recognized influence on decision making, we have little understanding about how consumers process information visually. Extant studies focus on numerical values and follow the traditional expected utility approach, using rating distributions as the basis of the probabilistic distribution that consumers use to make inferences and maximize consumption utility (Sun,

[1]IMDB (`https://www.imdb.com/`) is one of the world's most famous movie-rating websites. Amazon, the world's largest online retailer, shows a similar distribution chart.

2012; Kuksov and Xie, 2010). Such works signify that consumers read histograms of ratings as a distribution, leading to the inclusion of a variety of statistical measures, such as the mean, variance, skewness, kurtosis, and even entropy and the Herfindahl-Hirschman Index (e.g., Häubl et al., 2000; Dellarocas, 2003; Dellarocas et al., 2007; Li and Hitt, 2008; Rosario et al., 2016). However, these numbers may not be the actual input that consumers rely on in their decision-making, and there are known cognitive biases in the interpretation of graphic displays in general and of histograms in specific (e.g., Graham, 1937; Zacks and Tversky, 1999; Lem et al., 2014; Boels et al., 2019). In other words, there is a gap between how consumers mentally process and compare distribution charts and how statistical measures are used in empirical models. More research is needed to understand how consumers interpret and utilize online ratings (Simonson, 2016).

We study consumer information processing based on distribution charts in the form of histograms, with experiments designed to study the outcome of visual decision-making. Our results demonstrate that the widely accepted mean-variance trade-off is only partially correct in determining consumer choice with distribution charts. In our experiments, consumers consistently make choices that contradict the predictions of the mean-variance framework. We attribute the biased decision to a visual distortion in perceived distribution. Salient components in these charts may mislead consumers and distort their perceptions of the rating distribution (e.g. Parkhurst et al., 2002; Torralba et al., 2006). A simple stylized model helps reveal that such distortion could result in mis-judgement about mean ordering. In other words, there is a first-order distortion resulting from visual decision-making. We name this distortion the Visual Distortion Bias (VDB). In our setting, the distortion refers to the difference between the actual mean rating and the perceived mean rating.

We design a series of experiments to detect and quantify the impact of the VDB on consumer choices. Specifically, our study yields the following main findings. First, we find that consumers are sensitive to the shape of distributions. Controlling the mean, top-heavy distributions (i.e., distributions with ratings clustered above the mean) are preferred over middle-heavy ones, and middle-heavy distributions are considered more desirable than bottom-heavy ones. The effect is robust with respect to the level of the mean. That is, individuals would rather choose a product with a high VDB, even if its mean rating is lower. Second, the VDB leads to deviations from the predicted behavioral patterns in classical frameworks of decision-making. We present counter-examples against the branch independence assumption and show even first-order stochastic dominance fail to hold in some cases. It is worth noting that the VDB as a visual distortion is orthogonal to existing theories about behavioral biases, such as the prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992). Third, we show that while our study is motivated by the setting of displaying consumer ratings, the same effect can be observed in more traditional settings of decision-making under uncertainty (e.g., lottery) when probabilities are displayed graphically.

Unlike previous studies of online product reviews, which mostly examine the "information content" of ratings and reviews, our study focuses on consumers' perceptions of the "shapes" of rating distributions. Such visual decision-making may distort consumers' perception of the truth that these charts aim to convey. We thus contribute to the literature by extending the discussion to visual information processing and the impact of graphical information presentations. Given our heavy reliance on ratings and the movement towards simple user interfaces on mobile devices, this investigation is not only theoretically interesting but also practically important. Furthermore, most existing studies in behavioral economics focus on people's perceptions about outcomes (e.g., the prospect theory); and ours is one of the first to look at a cognitive bias rooted in visual distortion on the frequency of the outcomes. While the prospect theory finds that individuals put higher weights on low-probability events, our results show that individuals perceive high-probability events to have even higher probabilities of happening. Last but not least, the study contributes to the research on misinterpretation of graphical presentations of data by proposing, identifying and quantitatively measuring the visual distortion bias.

In the following, we first review the literature. We then analyze an illustrative model of consumers' tendency to amplify the visual distortion of distributions. Based on the model, we report the design and results of a series of experiments. We conclude the paper with a discussion about the implications for practice and future research.

2 Literature

Websites and Internet platforms widely adopt consumer rating systems. In online rating systems, consumers voluntarily and openly contribute ratings and text reviews for products and services. The creation of online ratings is motivated by self-selection, social influence, and strategic manipulation. First, online reviews suffer from the acquisition bias and the under-reporting bias, which results in a J-shaped distribution of online product ratings (e.g., Hu et al., 2009; Godes and Silva, 2012; Hu et al., 2017). Consumers are strategic in choosing which products to review and what ratings to give (Shen et al., 2015), and the interactions between sellers and buyers can also result in reporting bias in online rating systems (Dellarocas, 2006; Ye et al., 2014). Second, social connections and social networks embedded in social media platforms affect the characteristics of user-generated content (Huang et al., 2017). As users get more attention, they alter rating contribution patterns as a result of the popularity effect (Goes et al., 2014). Wang et al. (2018) identify significant social influence in the generation of online product ratings with a quasi-experimental design. Third, online ratings may also suffer from manipulation. Mayzlin et al. (2014) provide evidence that firms manipulate online reviews in response to competition. Luca and Zervas (2016) examine the use of fake reviews by restaurants. Overall, research on the generation of online reviews suggests that ratings are often biased signals of product quality. Nevertheless, consumers rely heavily on user-generated ratings in making purchasing decisions without considering these biases (De Langhe et al., 2016).

There is plenty of evidence that online ratings influence product sales in a variety of e-commerce contexts (e.g., Ba and Pavlou, 2002; Godes and Mayzlin, 2004; Chevalier and

Mayzlin, 2006; Liu, 2006; Dellarocas et al., 2007; Duan et al., 2008; Forman et al., 2008; Zhu and Zhang, 2010; Moe and Trusov, 2011; Ho-Dac et al., 2013). Early studies confirm the causal impact of rating volume (Duan et al., 2008; Liu, 2006; Xiong and Bharadwaj, 2014; Gu et al., 2012) and rating valence on product sales (Dellarocas et al., 2007; Zhu and Zhang, 2010; Chintagunta et al., 2010). Later studies extend the discussion to investigate the market impact of negative ratings (Chevalier and Mayzlin, 2006; Hiura et al., 2009), variance of ratings (Sun, 2012), dynamics of ratings (Godes and Silva, 2012), multi-dimensional ratings (Archak et al., 2011; Chen et al., 2017), and metrics such as skewness and kurtosis (e.g., Rosario et al., 2016). These studies generate important insights regarding how online ratings influence product sales. However, the literature is inconclusive about which particular metrics drive the effects (Rosario et al., 2016).

Consumers often face challenges in using all the information provided by online rating platforms (DellaVigna and Pollet, 2009). Research on how consumers make use of online ratings in their choices primarily focuses on how features of review content influence consumer perception (e.g., Mudambi and Schuff, 2010). It have been shown that negative reviews tend to be voted as more helpful than positive ones (e.g., Sen and Lerman, 2007; Cao et al., 2011; Chen and Lurie, 2013), and the usefulness of negative reviews is moderated by the confirmation bias (Yin et al., 2016). There is experimental evidence that the figurativeness of review content (Kronrod and Danziger, 2013), emotions embedded in online reviews (Yin et al., 2014; Zhu et al., 2014), and explanation type (action focus versus reaction focus) (Moore, 2015) can influence consumer perception. Text mining techniques are also useful in understanding how content features influence consumer perception. For example, Ghose and Ipeirotis (2011) mine the content of online reviews to identify influential text-based features and analyze their economic impact. While it is important to understand the making of helpful and thus influential reviews, we also need to find effective ways to present information embedded in online ratings data.

Human perception of data depends not only on the content but also on the presentation format (e.g., Chetty et al., 2009). Humans have developed great visual skills, such as the skill to detect edges and discontinuity, things that stand out, and variations in color, shape, and motion; to recognize patterns; and to retrieve information using visual cues (Kosslyn, 1994). Graphic displays, such as histograms and line-charts, are widely adopted to convey statistical information and facilitate inference. As more information becomes digital, a large number of visualization tools has been created to help decision makers. It has been shown that graphics are more effective than numerical values in conveying risk information and discouraging risk-taking behavior (Stone et al., 1997). However, graphical presentations can lead to biased interpretations and result in decision biases (Cleveland and McGill, 1984, 1985; Raghubir and Krishna, 1999; Krider et al., 2001; Lurie and Mason, 2007). For example, Spence (1990) show that judgement error depends on the graphical elements used to present the data. Salient features in a graph attract disproportionately more attention (e.g. Parkhurst et al., 2002; Torralba et al., 2006). Individuals overestimate the relative frequency or probability of more vivid information (Sherman et al., 1985).

Various misinterpretations of histograms have been described in the literature (e.g., Lem et al., 2014; Boels et al., 2019). Graham (1937) find that features such as axis orientation, coarseness of scale units, width of bars affect individuals' processing of data. Individuals are more inclined to interpret data presented in bar charts as discrete data point comparisons, while they interpret data presented in lines as trends (e.g., increasing, decreasing) (Zacks and Tversky, 1999). Further, Newman and Scholl (2012) find that people judge points that fall within a bar in a bar-chart as being more likely than points equidistant from the mean, but outside the bar as if the bar contained relevant data. Misinterpretation of graphically presented data may lead to critical choice biases.

Recent studies shed lights on how consumers may misinterpret distributional information presented as histograms. Luca and Smith (2013) document situations where consumers rely on very coarse information while ignoring finer details. He and Bond (2015) propose that consumers' interpretation of online rating dispersion depends on the extent to which tastes in a product domain are perceived to be dissimilar. Using experimental studies, they demonstrate that participants presented with online rating distributions were more tolerant of dispersion in taste-dissimilar product domains than taste-similar product domains, and the difference was driven by underlying attributions. Only recently have researchers started to examine graphical presentations of online ratings. Hu et al. (2017) study how consumers interpret polarized ratings from the perspective of self-expression needs. Fisher et al. (2018) propose that consumers exhibit a binary bias in interpreting user ratings. Despite increased research interests, there lacks research that (1) examines how the interpretation of distribution charts may influence consumer choices, and (2) gives proper quantitative measures to the potential visual distortions in the online ratings literature.

In the current study, we examine the impact of consumers' visual processing of salient features when examining rating distribution charts. We focus on the decision bias arising from visual presentation of data and develop an illustrative model of visual decision making to derive a quantitative Visual Distortion Bias measure. We examine the impact of such a bias with experiments in the context of online ratings and show that the Visual Distortion Bias can lead to the violation of previously well-established decision rules.

3 Visual Distortion Bias

The left panel of Figure 1 shows the rating distribution of the movie *Ocean's Eight* on IMDB. The movie has an average rating of 6.3 and a variance of 5.92. About 25.2% of the reviewers gave seven stars to the movie, with six-star ratings coming in second, representing 19.5%. Exposed to such a histogram, consumers will form an evaluation about the movie based on its rating distribution. Previously, consumers, researchers and platforms believe that the histogram presented is the actual input for evaluation, and thus the mean and the dispersion of the presented distribution determines consumer choices. Recent studies start to challenge this assumption. For example, in a study of online ratings, Fisher et al. (2018)

find that, rather than accounting for each level of user ratings in forming the evaluation, consumers exhibit a binary bias in interpreting the ratings. In other words, within positive and negative bins, people do not sufficiently distinguish more extreme values (5s and 1s) from less extreme values (4s and 2s). Different from their study, our proposal is that consumer perception weighs more on the salient bars (irrespective of the rating level).

We next argue, with a highly stylized model, that the formation of product evaluation is subject to participants' visual processing of the rating distribution and it may be distorted.

3.1 An Illustration of Bias in Perceived Average Rating

Since consumers cannot precisely calculate the mean of ratings, they rely on a quick visual assessment of the distribution of ratings. Studies on visual cognition have found that salient features in a graph attract disproportionately more attention while less salient ones can get ignored (e.g. Parkhurst et al., 2002; Torralba et al., 2006). Visual focus on more salient components (or longer bars) in a histogram therefore should lead to over-weighting of the corresponding rating level. We use an illustrative model to demonstrate that consumers' processing of visually presented data may lead to biased perceptions.

To illustrate the impact of such probabilistic over-weighting, we assume that the perceived length of a rating level (i.e., the length of the bar in the chart) is a transformed function of the actual length, denoted by $t(p_i|p_{-i})$, where p_i is the actual length of rating i and p_{-i} represents the length of the other rating levels.

Since we do not know the form of the transformation function $t(\cdot)$, we apply Taylor expansion to obtain its approximation. We have

$$t(p_i) = t(0) + t'(0)p_i + \frac{1}{2}t''(0)p_i^2 + R_2(p_i)$$

where $R_2(p_i)$ is the higher-order residual term. We can then write the transformation as a quadratic function of the original probabilities for illustration purpose:[2]

$$t(p_i) = p_i + \lambda p_i^2,$$

where λ, a function of $t'(0)$ and $t''(0)$, is a curvature measure that captures the visual distortion. When $\lambda > 0$, the longer bars in the histogram are over-weighted.

We normalize the transformed lengths of the rating levels so that the probabilities add up to 1 and are well-defined:

$$w(p_i|p_{-i}) = \frac{t(p_i)}{\sum_j t(p_j)} \tag{1}$$

To illustrate the distortion, Figure 2 shows a "perceived distribution" for the movie *Ocean's Eight* when $\lambda = 5$. As the figure shows, the probability distortion resulting from the visual focus on the salient bars leads to the "perceived distribution", which has a higher mean ($\mu = 6.47$) and a lower variance ($\sigma^2 = 4.86$) compared to the actual distribution.[3]

3.2 Measure of Visual Distortion Bias

Formally, the perceived average rating, $\overline{x}_s$, is represented by Equation 2. In the calculation, the perceived probability of a rating level depends on both its actual probability and the probability of other levels.

$$\overline{x}_s = \sum_i w(p_i|p_{-i}) \cdot x_i = \sum_i \frac{p_i + \lambda p_i^2}{\sum_j p_j + \lambda p_j^2} x_i \tag{2}$$

We define the difference between the subjective and the actual average ratings as the Visual Distortion Bias (or VDB):

[2]To be more precise, we present an analysis of the impact of the residual term on VDB in Appendix A. According to the analysis, it is reasonable to focus on the lower order terms in the calculation.

[3]It should be noted that we do not assume that consumers actually refer to the distorted distribution in making decisions. This figure simply illustrates the impact of the distorted perception.

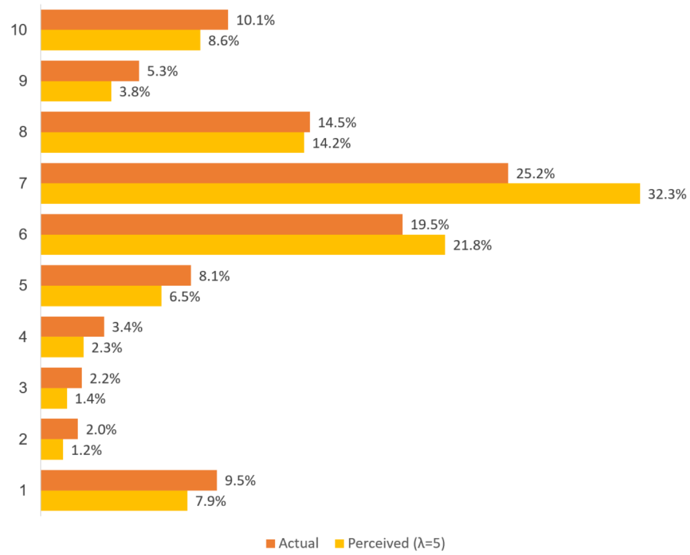

Figure 2: Illustration of Distribution Distortion

$$\text{VDB} = \overline{x}_s - \overline{x} = \frac{\lambda}{1 + \lambda \sum_j p_j^2} \sum_i p_i^2 (x_i - \overline{x}). \tag{3}$$

We can empirically calibrate the scaling factor involving λ. To facilitate experimental design, we define the baseline VDB as follows:[4]

$$\text{VDB}_{base} = \sum_i p_i^2 (x_i - \overline{x}). \tag{4}$$

We can compare the VDB with the average rating. To calculate the average rating ($\bar{x} = \sum p_i \cdot x_i$), the rating levels ($x_i$) and the frequencies ($p_i$) enter the equation equally. However, in VDB_{base}, the quadratic form of the frequencies (p_i^2) and the deviations from the mean ($x_i - \bar{x}$) together influence the perception of rating distributions.

[4]In the experiments, we calculate the baseline VDB as a reference to help us predict the outcomes. With experimental data, we are able to estimate the curvature, λ, empirically. Details are available upon request.

3.3 Discussions

A few observations can be made regarding the illustrative model. First, the VDB represents a first-order distortion in the perceived mean rating. It depends on both the viewer's focus on the longer bars, as captured by the variable λ, and on the shape of the distribution, as captured by the VDB_{base} variable. Second, VDB_{base} resembles the calculation of skewness. In other words, if consumers exhibit bias resulting from visual focus ($\lambda > 0$), we should observe a consumer preference for positively skewed distributions. Third, as the VDB results from a subjective distortion of the probability, it may overturn the probability order between objective distributions. That is, a movie with a higher mean rating may be less desirable than one with a lower rating, depending on the shape (skewness) of the distribution. In the following section, we design a series of experiments to test the impact of the VDB on consumer choice.

The VDB is a type of distortion that results from the "shape" of the distribution. Previous studies in the behavioral economics and marketing literature also identify distortions in perceived probabilities. Weighting function in the prospect theory is non-linear, implying a distortion of the objective probability. In the original prospect theory (Kahneman and Tversky, 1979), decision makers overweight extreme outcomes. In the cumulative prospect theory, Tversky and Kahneman (1992) propose a rank-dependent weighting distortion. The subjective probability distribution depends on the outcomes rather than the relative probability levels of the outcomes. In the marketing literature, Fisher et al. (2018) propose a binary bias in interpreting online ratings. They find that people prefer top-heavy rating distributions. Their binary bias, however, focuses on value non-linearity in rating interpretations. Similar to the prospect theory, the binary bias distortion arises with respect to the values (levels of ratings). In this paper, distortion results directly from the probability distribution rather than from the outcomes (i.e., $w(p_i|p_j)$ only depends on the probability distribution). This distortion is a result of visual decision-making.[5]

[5]Due to fundamental differences in the underlying mechanisms between our model and the previous ones,

The illustration and discussion presented in this section is informative for the design of the experiments. First, it suggests that a decision bias may arise from a visual distortion of the actual distribution of ratings. The distortion may change the perceived mean rating. Second, the illustration generates a baseline measure of the distortion (the baseline VDB) as a feature of the rating distribution. We also show that there is a trade-off between the mean and the distortion factor. In other words, the VDB is a first order distortion that may results in violation of well-established decision patterns under the mean-variance framework. Finally, it helps to focus our attention on the basic properties of the distribution rather than the utility that is associated with the outcome. Since the distortion is a basic property of human perception of histograms, we expect generalization of the research findings to other contexts of decision-making based on visual assessment of histograms.

4 Experimental Design and Results

To examine consumer decision-making assisted by a distribution graph of online ratings, we conduct a series of experimental studies on Amazon Mechanical Turk. We require the subjects to be from 18 to 60 years' old, living in the United States, having a Human Intelligence Task (HIT) approval ratio of above 95%, and restrict them to participate in the experiment only once. Each study examines the impact of the VDB from a unique perspective. These experiments enable us to control for or eliminate other decision factors present in field settings (e.g., product/service content, pictures, text reviews, etc.). Most studies (Studies 1 to 7) follow the same procedure as described below.

Participants were asked to imagine that they were to choose, based on user ratings, a movie to watch from two alternatives (i.e., movie P and movie Q). While we expect the VDB to be present in broad choice scenarios, we choose to contextualize the experiments as movie choices because consumers routinely refer to online reviews when selecting movies to watch,

our model generates interesting and previously unreported effects that are supported by our experimental results. We provide a comparison between our theory and previous theories about decision biases in the appendix.

thus participants are familiar with the context (Liu, 2006; Chintagunta et al., 2010). Online movie ratings play a particularly important role in providing information to consumers (Moe and Trusov, 2011; Rosario et al., 2016).[6]

Presentation of the user ratings resembled the 10-star histograms on IMDB (Figure 1).[7] In each study, participants made one or more choices sequentially depending on the design of the particular study (for example, Study 1 included two comparisons and Study 2 contained only one comparison). Participants needed to make a choice in each comparison before they could move to the next one. There was no time limit, and the order of the pair of choices (left vs. right) was randomized in each study.

4.1 Study 1: The Impact of the VDB

Study 1 aims to demonstrate the effect of base VDB while controlling the mean and variance of the review distribution. 101 participants were recruited for Study 1. Each participant saw two pairs of movie ratings (see Figure 3, 4). Study 1-1 has a unimodal distribution, with a mean of 7.0 and a variance of 1.2 for both movies, and Study 1-2 has a bi-modal distribution, with a mean of 7.0 and a variance of 2.0 for both movies. The only difference in each pair is their base VDB score.

The result shows that although the two movies had the same mean and variance, participants significantly preferred the movie with a higher VDB. Specifically, 70 out of 101 participants ($p < 0.001$) in Study 1-1 and 66 out of 101 participants ($p = 0.003$) in Study 1-2 preferred the movie with the higher VDB.

4.2 Study 2: Dominated Effect of Variance

Since the VDB is a first-order effect, its impact should be able to dominate the effect of variance. Study 2 aims to show that the effect of variance in consumer decision-making

[6] Although we selected the setting of user ratings in this study, the VDB effect should be present whenever histograms are displayed. We examine this possibility in Study 7.

[7] Presented under “Experimental Stimuli” in the Figures.

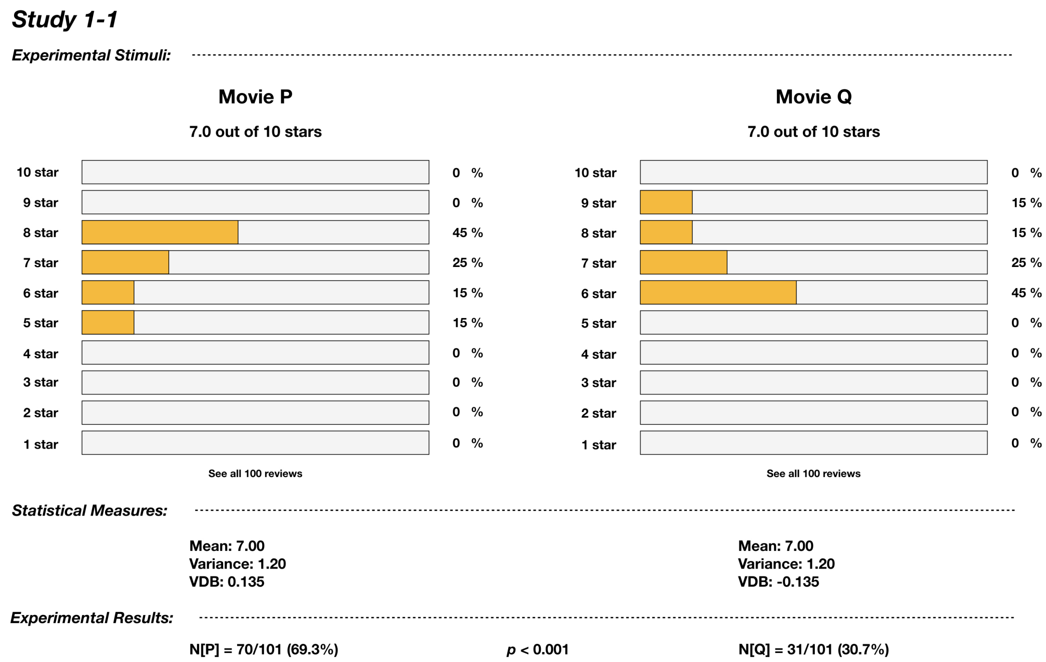

Figure 3: Study 1-1

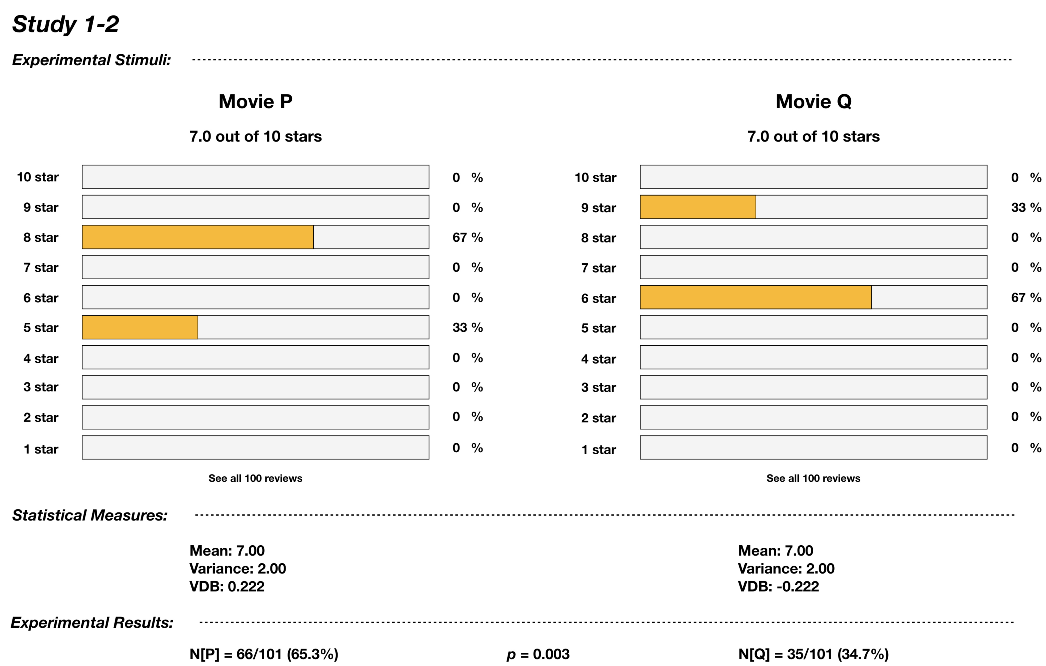

Figure 4: Study 1-2

could be dominated by that of the VDB. Each of the 104 participants in the study saw a pair of movie ratings (see Figure 5). Both movie ratings (movie P vs. Q) have the same mean of 7.0, where movie P has a lower variance as well as a lower VDB score. We observe that 64 out of 104 ($p = 0.024$) participants chose the movie with a higher VDB and a higher variance, contradicting the traditional mean-variance prediction.

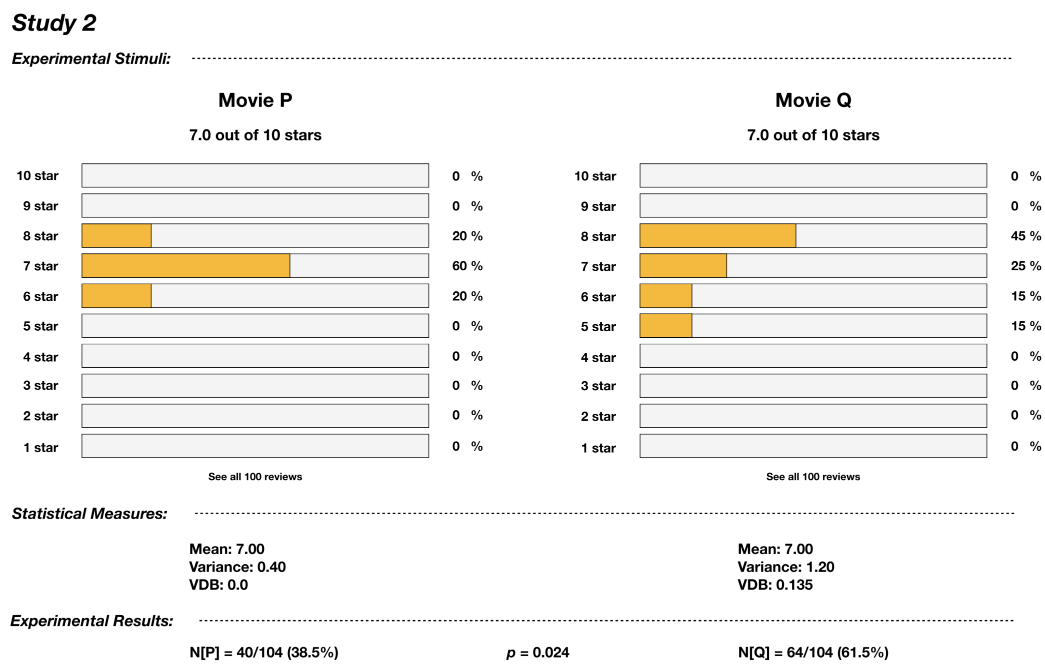

Figure 5: Study 2

4.3 Study 3: Mode Position

Studies 1 and 2 demonstrate that consumers prefer rating distributions with a higher VDB, and the effect can dominate the effect of the variance. One could argue that the findings could be explained by preference over the mode of distributions (i.e., the longest bar in the histograms). Study 3 aims to show that the VDB could predict consumers' preference when the mode of the distributions are the same, and thus rule out this interpretation.

Each of the 202 participants in Study 3 saw two pairs of movie ratings (see Figures 6, 7).

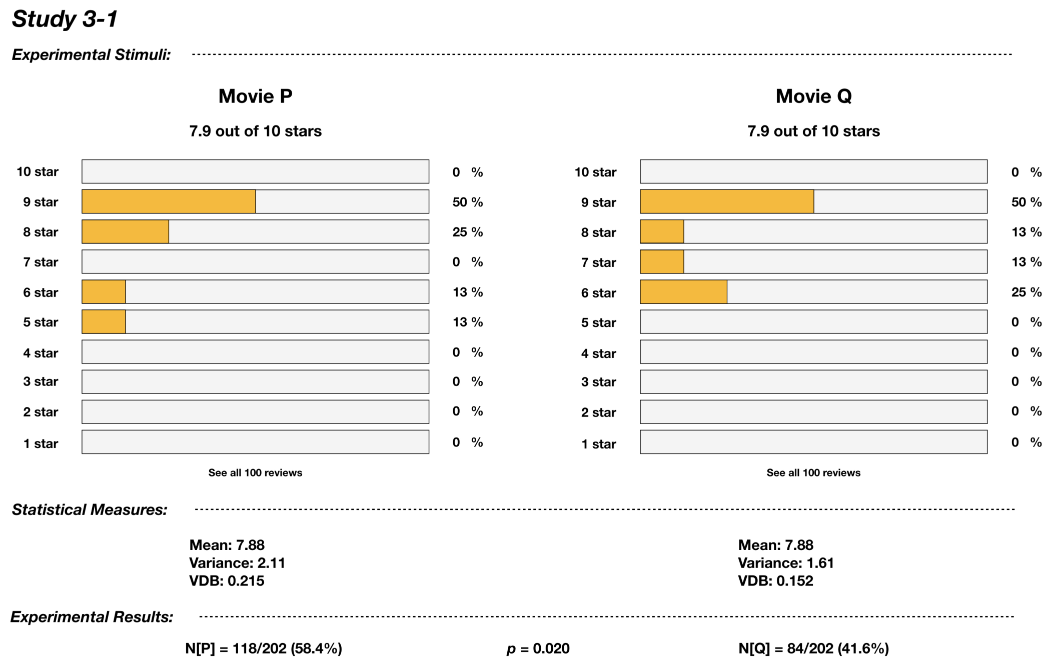

Figure 6: Study 3-1

Each pair of movies shares the same mean and the same mode. Yet, movie P has a higher VDB than that of movie Q. If the mode is the only factor that determines product choices, participants should be indifferent between these two movie options in both pairs. However, 118 out of 202 ($p = 0.020$) participants chose movie P in the first pair (Study 3-1), and 130 out of 202 ($p < 0.001$) participants chose movie P in the second pair (Study 3-2).

4.4 Study 4: Branch Independence

Study 4 aims to show that the VDB could result in violation of the branch independence assumption. Branch independence is a weaker assumption than Savage's independence axiom and states that if two random events have a common outcome for an event of known probability, the value of that common outcome should have no effect on the preference order induced by other probability-outcome branches (Birnbaum and McIntosh, 1996). In Study 4-1, both distribution charts have the same component at 9 stars with a proportion of 35%.

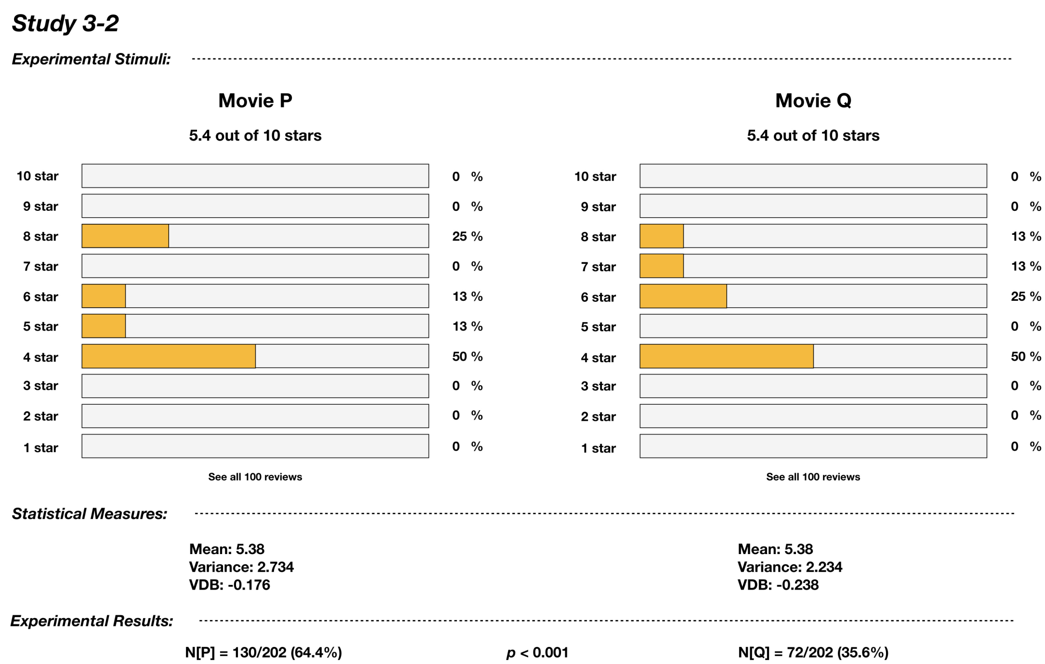

Figure 7: Study 3-2

Movie P has 5 rating bars uniformly distributed from 8 to 4 stars, and movie Q has all the remaining 65% of ratings located at 6 stars. The means of the two distributions are the same (7.1). In Study 4-2, we move the common components from 9 stars to 3 stars and keep all the other bars intact (see Figure 9). According to the branch independence assumption, consumers' choices should be consistent across Study 4-1 and Study 4-2. In other words, if one prefers movie P in Study 4-1, they should also prefer movie P in Study 4-2. Our model, however, predicts that participants will prefer movie P in Study 4-1 but movie Q in Study 4-2 (i.e., movies with higher VDB scores).

202 participants participated in Studies 4-1 and 4-2. As predicted, 125 out of 202 participants ($p < 0.001$) chose movie P in Study 4-1, while 127 out of 202 participants ($p < 0.001$) chose movie Q in Study 4-2. These results show that the participants did not have a consistent preference, as the branch independence assumption indicated, but instead they behaved mostly according to our prediction.

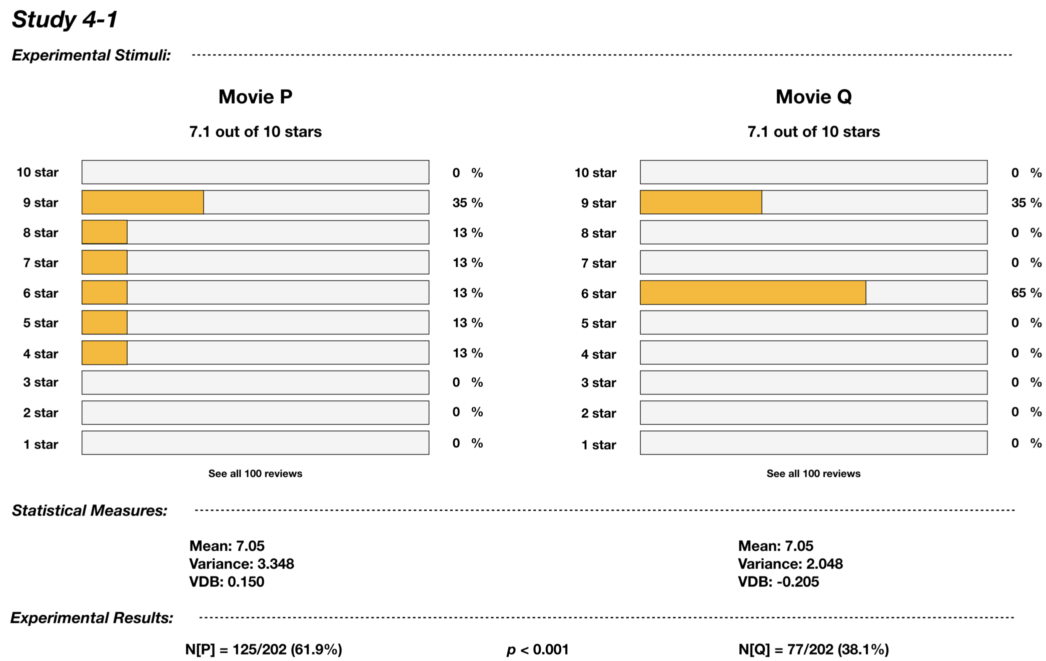

Figure 8: Study 4-1

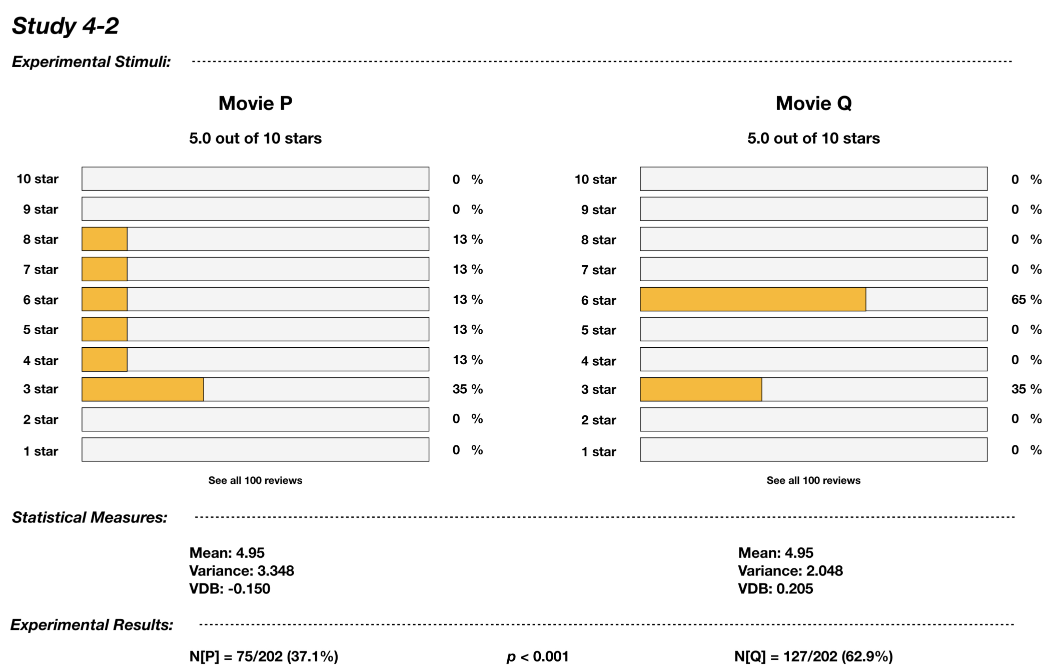

Figure 9: Study 4-2

4.5 Study 5: First-Order Statistical Dominance

One of the most well-established choice patterns under uncertainty is first-order stochastic dominance (FOSD). A distribution F is said to be first-order stochastic dominated by another distribution G when the cumulative distribution function (CDF) of F is greater than that of G everywhere on the support set. FOSD implies that a choice with a dominated distribution should never be preferred. In Study 5, we aim to show that even FOSD can be challenged by the VDB. In other words, as a result of the VDB, individuals would violate FOSD and make choices that are strictly inferior.

The 302 participants in Study 5 were asked to choose between a pair of movies where the rating distribution of movie P is dominated by the distribution of movie Q by FOSD and has a higher VDB than that of movie Q (see Figures 10 and 11). The results show that the majority of participants chose movie P over movie Q ($p = 0.03$).[8] Consistent with our prediction, even though movie Q had a rating distribution that first-order stochastic dominates movie P, participants still preferred movie P, which is a violation of FOSD. This study also demonstrates the importance of recognizing the VDB in decision-making. To the best of our knowledge, violations of FOSD has not been reported before. None of the decision biases from the distortions of the value function in the literature could explain a FOSD violation.

4.6 Study 6: The Trade-off between the Mean and the VDB

So far, we have demonstrated that the effect of the VDB exists, dominates the effect of variance, goes beyond preference over mode, and results in violations of well-established decision rules such as branch independence and first order stochastic dominance. In Study

[8] Here, we report the p-value for a null hypothesis that individuals chose movies P and Q with equal probability. If we used rational decision-making as the baseline model, the null hypothesis would be that all individuals should choose movie Q. (Different from previous movie pairs that have a 50-50 divide as the benchmark, in the case of FOSD, the benchmark is 0-100. That is, all participants should unconditionally choose movie Q.) With the 0-100 benchmark, the p-value is even smaller and the null hypothesis will be rejected with even higher statistical confidence.

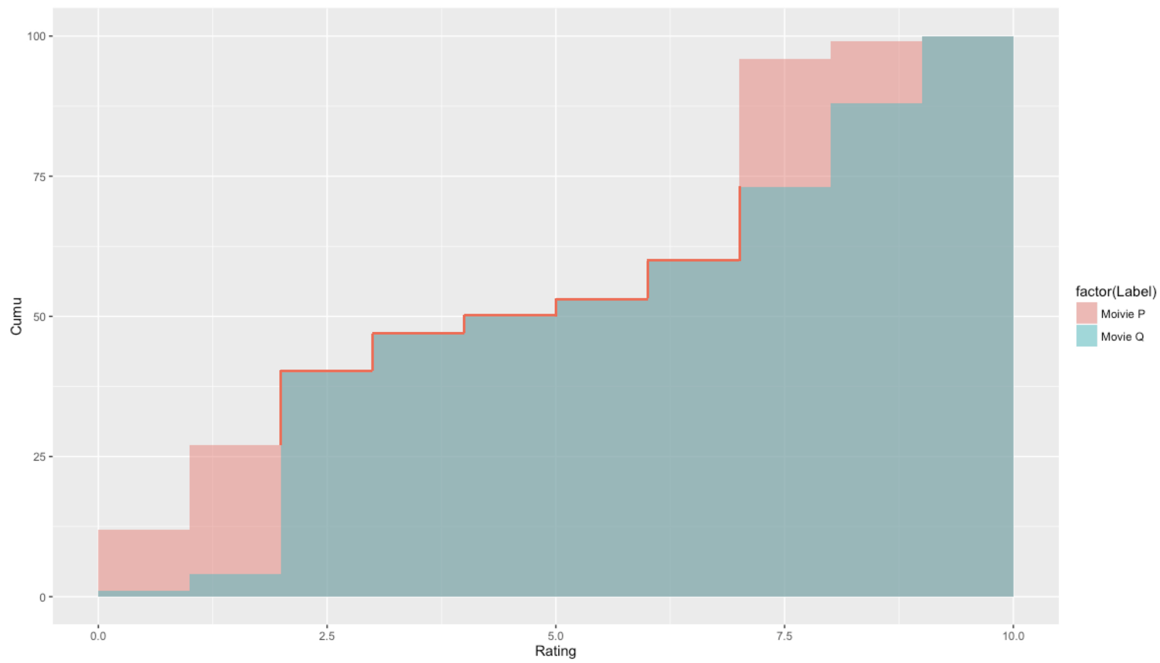

Figure 10: Cumulative Distribution Function of Stimuli in Study 5

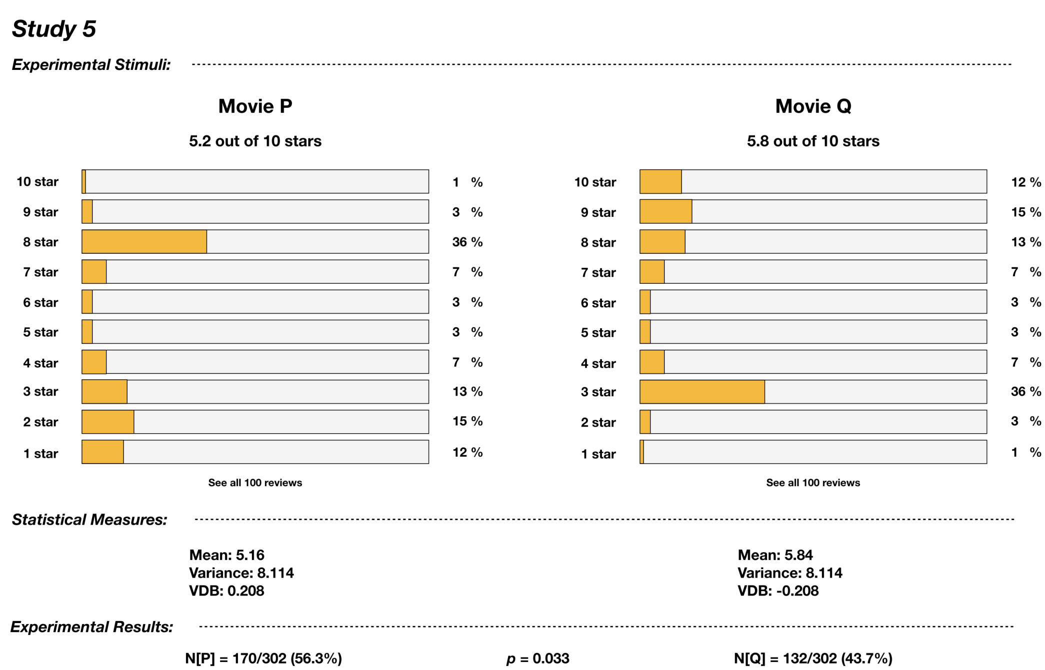

Figure 11: Study 5

6, we design a series of comparisons to detect the trade-off between the mean and the VDB, as both are first-order decision factors and both have substantial impacts on consumer choices. With this study, we would like to establish a measure of relative importance between the two. Sixteen pairs of comparisons were implemented.[9]

We design the study by varying the levels of three factors: rating range (high vs. low), variance (high vs. low), and differences in the means (0, 0.1, 0.2, 0.4). In each pair, movie P has a positive base VDB, and movie Q has a negative base VDB. When the two charts in each pair do not have equal means, movie P always has a smaller mean so that we can examine the trade-off between the mean and the VDB. Given the finding that movie P is preferred due to the VDB, we gradually reduce its mean to make movie Q more attractive. This way we can examine the trade-off between the mean and the VDB. Table 1 summarizes the design. Figure 12 presents the results in graph form.

Table 1: Results of Study 6

	Mean			Variance		VDB			Choice		
No.	P	Q	Diff.	P	Q	P	Q	Diff.	P	Q	Sig.
1	7.0	7.0	0.0	1.20	1.20	0.135	-0.135	0.270	70	30	***
2	7.0	7.1	−0.1	1.23	1.23	0.123	-0.123	0.246	58	40	*
3	6.9	7.1	−0.2	1.25	1.25	0.110	-0.110	0.220	44	56	
4	6.8	7.2	−0.4	1.28	1.28	0.083	-0.083	0.166	38	62	**
5	7.0	7.0	0.0	2.00	2.00	0.222	-0.222	0.444	63	37	**
6	7.0	7.1	−0.1	2.05	2.05	0.205	-0.205	0.410	65	36	***
7	6.9	7.1	−0.2	2.10	2.10	0.182	-0.182	0.364	58	42	
8	6.8	7.2	−0.4	2.16	2.16	0.144	-0.144	0.288	37	63	**
9	4.0	4.0	0.0	1.20	1.20	0.135	-0.135	0.270	74	26	***
10	4.0	4.1	−0.1	1.23	1.23	0.123	-0.123	0.246	66	36	***
11	3.9	4.1	−0.2	1.25	1.25	0.110	-0.110	0.220	58	42	
12	3.8	4.2	−0.4	1.28	1.28	0.083	-0.083	0.166	37	63	**
13	4.0	4.0	0.0	2.00	2.00	0.222	-0.222	0.444	78	22	***
14	4.0	4.1	−0.1	2.05	2.05	0.205	-0.205	0.410	65	36	***
15	3.9	4.1	−0.2	2.10	2.10	0.182	-0.182	0.364	59	41	*
16	3.8	4.2	−0.4	2.16	2.16	0.144	-0.144	0.288	48	52	

Note: *p<0.1; **p<0.05; ***p<0.01

First, in pairs with equal means (Studies 6-1, 6-5, 6-9, and 6-13), rating distributions

[9] The 16 pairs of comparisons are presented in Appendix C.

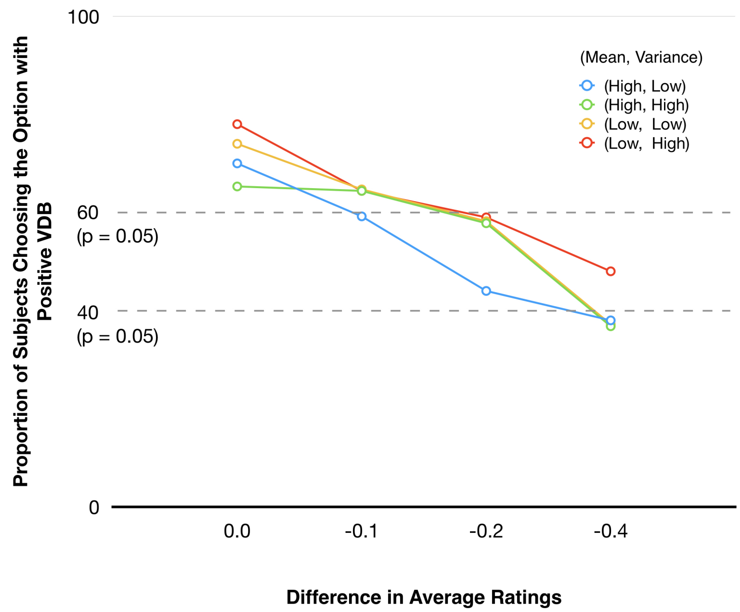

Figure 12: Summary of Results of Study 6

with positive VDB are always preferred, confirming the positive impact of the VDB. Second, regarding comparisons between the pairs with a 0.1 difference in means (Studies 6-2, 6-6, 6-10, and 6-14), three pairs of comparisons show significant results that ratings with a positive VDB are preferred in spite of a slightly lower mean. Third, regarding comparisons between the pairs with a 0.2 difference in means (Studies 6-3, 6-6, 6-9, and 6-12), the tension between the mean and the VDB reached the balance and the participants showed equal preference between the two movies. Finally, when the difference in means is higher at the level of 0.4 (Studies 6-4, 6-8, 6-12, and 6-16), distributions with higher means are more preferred. Roughly, the mean needs to be 0.4-star higher in order to compensate for a small difference (about 0.166 $\sim$ 0.288) in VDB. While we cannot enumerate all possible combinations, this study shows that there is a trade-off between the mean and the VDB. The relative importance of the VDB is higher when the level of the mean is low and the variance is high.

4.7 Study 7: Lottery Setting

So far we use movie ratings as a setting to examine the VDB. As the discussion in Section 3 suggests, similar distortions should be present in other contexts that involve visual comparisons of distributions. In Study 7, we try to detect the VDB in a lottery setting.

The 100 participants in Study 7 were told that they could participate in one of two lotteries that each had an average payoff of \$7. The probabilities of each possible payoff amount were displayed in a distribution chart (see Figure 13). We adopted the same stimuli as in Study 1-1. Participants were asked to indicate in which lottery they would like to participate for free.

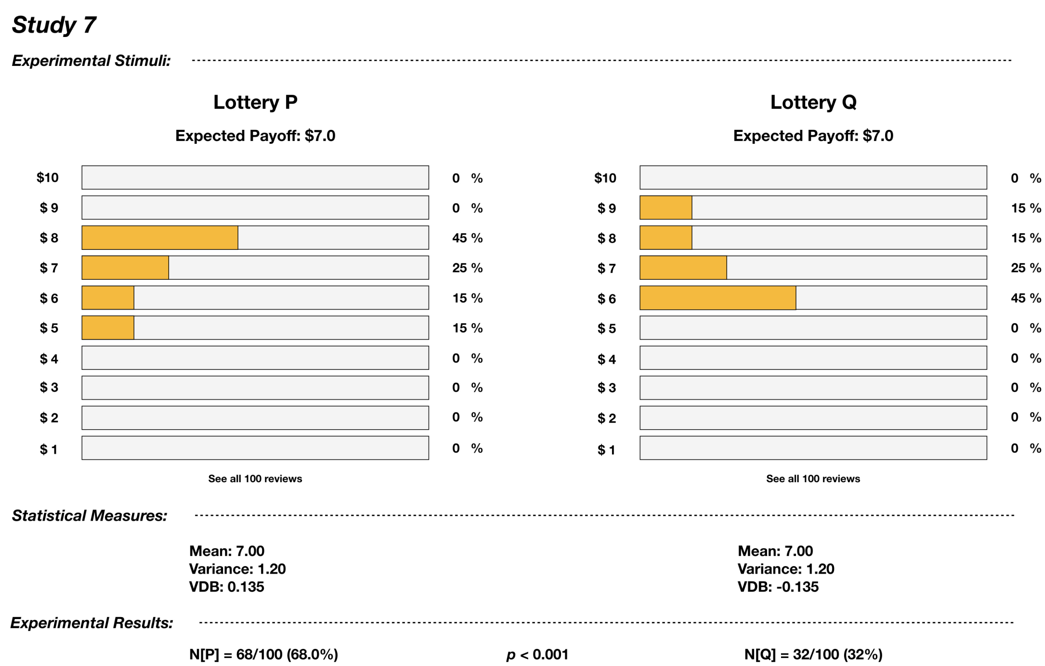

Figure 13: Study 7

Out of the 100 participants, 68 ($p < 0.001$) chose the lottery whose VDB was higher despite the two lotteries having the same mean and variance. This result is consistent with the result of Study 1-1. In other words, the effect of the VDB can be generalized to other contexts.

4.8 Study 8: 5-star Setting

In the studies reported so far, we adopt a 10-star histogram display similar to that of IMDB. In practice, distribution presentations may not always have 10 levels of rating outcomes. For example, many online shopping websites use a 5-star distribution chart (e.g., Amazon). Study 8 aims to test the robustness of the VDB effect in a 5-star setting. One hundred participants were recruited for the study. The procedure was similar to that of the previous studies. Participants were asked to choose between a pair of 5-star reviews that both had a mean of 3.95 out of 5 (see Figure 14). Our results show that the movie with a higher VDB was preferred by 66 out of the 100 participants ($p = 0.002$).

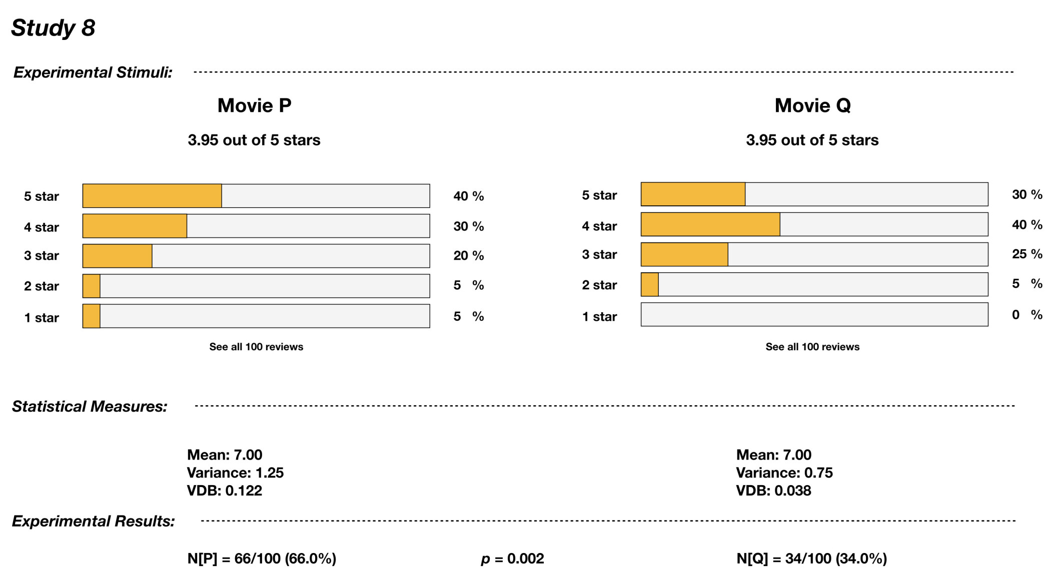

Figure 14: Study 8

4.9 Study 9: Perceived Mean of Ratings

In previous studies, participants face a pair of rating graphs and are asked to make a choice without stating their rationales. As our illustrative model suggests, the VDB is a bias with respect to the difference between the actual average rating and the perceived average rating. In this study, we aim to test whether the VDB indeed arises from distorted perceived average ratings. Participants were shown each of the two rating distributions sequentially with the order randomized. After seeing each distribution, they were asked to estimate the average rating.

We recruited 203 subjects in this study. Participants were asked to imagine that they were selecting movies online and were presented with a movie review without the average rating disclosed. Then they were asked to estimate the average rating of the movie. Each participant examined 2 movie rating distributions (Movie P in Figure 15 and Movie Q in Figure 16, with the order randomized) sequentially, and estimated an average rating for each movie independently. Movie P has an average rating of 5.2/10 and movie Q has an average rating of 5.8/10. Although movie Q has a higher average rating, it has a lower VDB. The results show that the majority (52.2%) of the participants (106 out of 203) assigned a higher score to movie P than movie Q, and 20% of the participants (39 out of 203) thought the two movies had the same mean. Overall, movie P has a perceived mean of 6.42 (SD=1.65) while movie Q has a perceived mean of 5.54 (SD=1.96). The results show that consumer perception of the average rating was indeed influenced by the VDB.

4.10 Real-World Rating Distributions

To shed light on the prevalence of the VDB in real-world rating distributions, we collect rating distribution data from the IMDB. Figure 17 shows the density of the VDB distribution across movie rating distributions on the IMDB. We can see that the distribution of VDB in real-world data is positively skewed. There are more movies with a positive VDB. Over 22% of the movies' absolute VDB level is larger than 0.1.

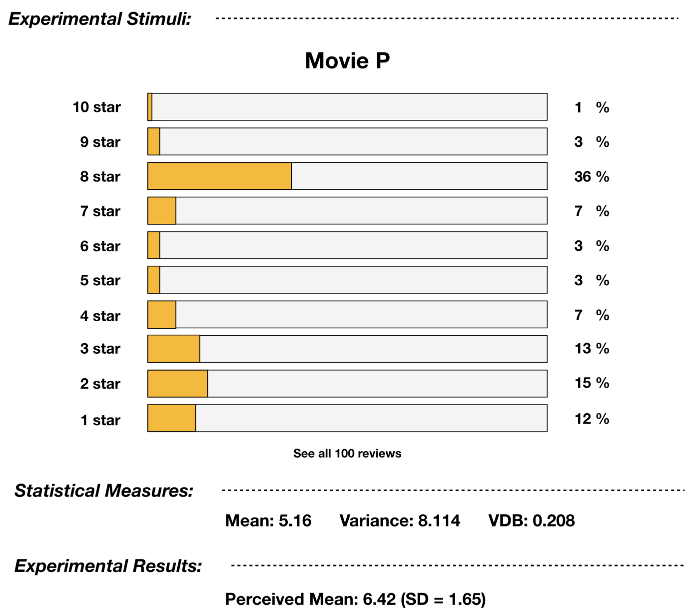

Figure 15: Study 9 Movie P

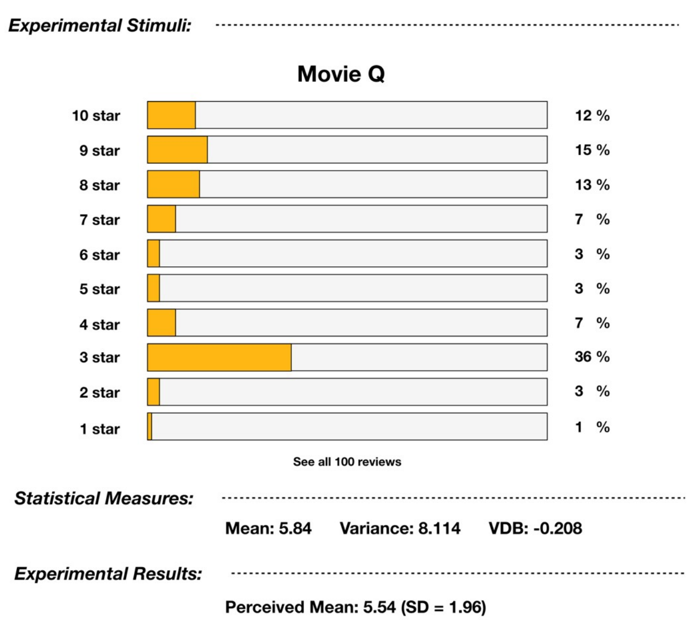

Figure 16: Study 9 Movie Q

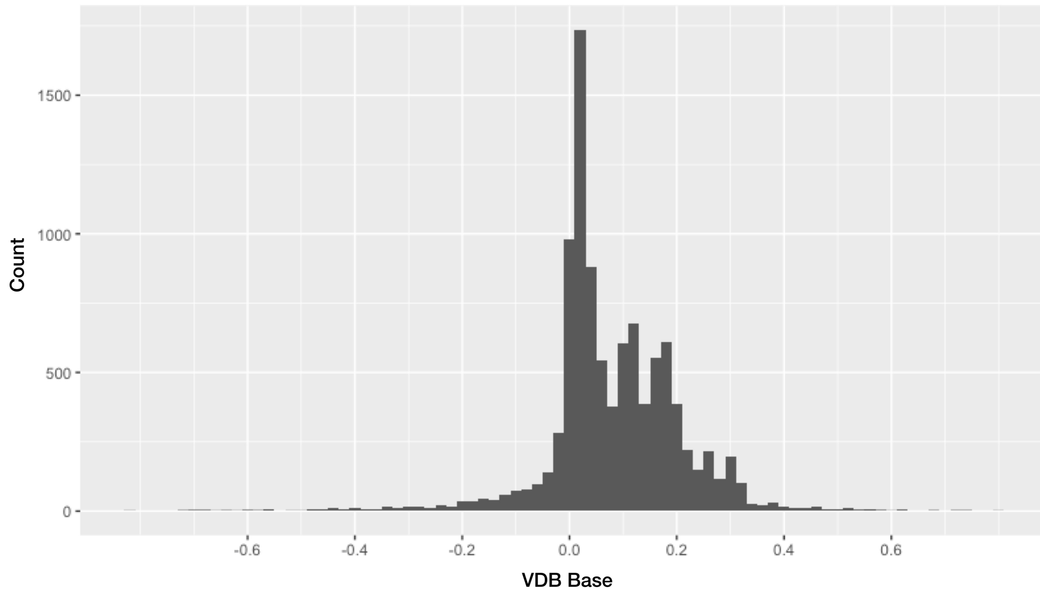

Figure 17: VDB Frequency in IMDB

5 Conclusion

We conduct a series of experiments in which participants choose between pairs of distributions of online ratings displayed as histograms. We find that the shapes of the distributions have a significant impact on consumers' perception of the mean ratings. Observed choices violate the predictions of the classical mean-variance framework of rational decision-making, including mean-variance trade off, branch independence, and first-order stochastic dominance.

Existing studies on online WOM often overlook how consumers utilize graphical decision aids that are commonly implemented by e-commerce and social media websites. Our study identifies a visual distortion bias that can lead to distorted and suboptimal consumer decision-making.

Theoretically, we propose an illustrative model and derive a measure of the bias: VDB. The model illustrates that graphically presented decision-making aids such as histograms may give consumers a distorted perception of the probability distribution of the bars in these histograms. The analysis leads to the discovery of a new first-order distortion in consumer decision-making that has not been documented in prior literature. The VDB dominates the effects of variance and plays a primary role in decision-making under situations when data are presented in histograms. Our study suggests that more research regarding graphical information presentation and visual biases should be conducted in online reporting systems. It has significant implications for marketing and system design.

As more big data-driven, graphically aided decision support systems become widely adopted in consumer markets and businesses, it is critical to deepen our understanding about how visual presentation of information influences decision-making in online environments.

This work has several limitations. First, we limit our discussions to the setting of online ratings. While we do show that the VDB extends to the lottery setting, it is both interesting and important to test the implications of the VDB in other decision contexts. Second, our analytical framework is a preliminary attempt to illustrate a decision bias resulting from visual distortions. It can be extended in many ways and will generate more theoretical predictions that can be tested in future empirical studies. Third, we only focus on the distortion to the perceived distribution without touching on consumer utility. It will be fruitful to integrate the illustrative model in an utility framework to understand its interactions with other decision biases, such as the prospect theory (Kahneman and Tversky, 1979) and the "binary bias" (Fisher et al., 2018). Such a framework will generate insights that allow us to compare models with different assumptions. Finally, to ensure the accessibility of the experiments, we asked the participants to compare between pairs of rating distributions. Although Study 9 suggests the bias is present in scenarios of standalone choices, visual decision aids may take many different forms in reality (e.g., Spence, 1990; Tversky, 1997). For example, it would be interesting to explore what might happen when individuals face more than two choices, whether other dashboard charts such as the pie chart may or may not have such distortion biases. To assess whether the identified impact of the visual distortion bias extends to standard histograms (i.e., those with vertical bars) or numerically presented data, we conducted additional exploratory studies and obtained similar findings.[10] This opens the door to many interesting research questions on different ways that human perception can deviate from the numerical values calculated by statistics. Future studies can extend our discussion and investigate the other types of distortions of data presentations.

[10]The detailed experimental design and results are available upon request. We thank anonymous reviewers for suggesting these tests.

References

Archak, N., Ghose, A., and Ipeirotis, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management Science*, 57(8):1485–1509.

Ba, S. and Pavlou, P. A. (2002). Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS quarterly*, pages 243–268.

Birnbaum, M. H. and McIntosh, W. R. (1996). Violations of branch independence in choices between gambles. *Organizational Behavior and Human Decision Processes*, 67(1):91–110.

Birnbaum, M. H. and Stegner, S. E. (1979). Source credibility in social judgment: Bias, expertise, and the judge's point of view. *Journal of Personality and Social Psychology*, 37(1):48.

Boels, L., Bakker, A., Van Dooren, W., and Drijvers, P. (2019). Conceptual difficulties when interpreting histograms: A review. *Educational Research Review*, 28:100291.

Cao, Q., Duan, W., and Gan, Q. (2011). Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach. *Decision Support Systems*, 50(2):511–521.

Chen, P.-Y., Hong, Y., and Liu, Y. (2017). The value of multidimensional rating systems: Evidence from a natural experiment and randomized experiments. *Management Science*, 64(10):4629–4647.

Chen, Z. and Lurie, N. H. (2013). Temporal contiguity and negativity bias in the impact of online Word of Mouth. *Journal of Marketing Research*, 50(4):463–476.

Chetty, R., Looney, A., and Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review*, 99(4):1145–77.

Chevalier, J. A. and Mayzlin, D. (2006). The effect of Word of Mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3):345–354.

Chintagunta, P. K., Gopinath, S., and Venkataraman, S. (2010). The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing science*, 29(5):944–957.

Cleveland, W. S. and McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association*, 79(387):531–554.

Cleveland, W. S. and McGill, R. (1985). Graphical perception and graphical methods for analyzing scientific data. *Science*, 229(4716):828–833.

De Langhe, B., Fernbach, P. M., and Lichtenstein, D. R. (2016). Navigating by the stars: Investigating the actual and perceived validity of online user ratings. *Journal of Consumer Research*, 42(6):817–833.

Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*, 49(10):1407–1424.

Dellarocas, C. (2006). Strategic manipulation of internet opinion forums: Implications for consumers and firms. *Management Science*, 52(10):1577–1593.

Dellarocas, C., Zhang, X. M., and Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, 21(4):23–45.

DellaVigna, S. and Pollet, J. M. (2009). Investor inattention and friday earnings announcements. *The Journal of Finance*, 64(2):709–749.

Duan, W., Gu, B., and Whinston, A. B. (2008). Do online reviews matter? – An empirical investigation of panel data. *Decision Support Systems*, 45(4):1007–1016.

Fisher, M., Newman, G. E., Dhar, R., and van Osselaer, S. (2018). Seeing stars: How the binary bias distorts the interpretation of customer ratings. *Journal of Consumer Research*, 45(3):471–489.

Forman, C., Ghose, A., and Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, 19(3):291–313.

Ghose, A. and Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, 23(10):1498–1512.

Godes, D. and Mayzlin, D. (2004). Using online conversations to study Word-of-Mouth communication. *Marketing Science*, 23(4):545–560.

Godes, D. and Silva, J. C. (2012). Sequential and temporal dynamics of online opinion. *Marketing Science*, 31(3):448–473.

Goes, P. B., Lin, M., and Au Yeung, C. M. (2014). Popularity effect in user-generated content: Evidence from online product reviews. *Information Systems Research*, 25(2):222–238.

Graham, J. L. (1937). Illusory trends in the observations of bar graphs. *Journal of Experimental Psychology*, 20(6):597.

Gu, B., Park, J., and Konana, P. (2012). Research notethe impact of external word-of-mouth sources on retailer sales of high-involvement products. *Information Systems Research*, 23(1):182–196.

Häubl, Gerald, and Trifts, V. (2000). Consumer decision making in online shopping environments: The effects of interactive decision aids. *Marketing Science*, 19(1):4–21.

He, S. X. and Bond, S. D. (2015). Why is the crowd divided? attribution for dispersion in online word of mouth. *Journal of Consumer Research*, 41(6):1509–1527.

Hiura, K., Kikumori, M., Kishimoto, K., Matsumoto, N., Nakagawa, M., Ujita, M., and Ono, A. (2009). Does negative e-wom affect consumer attitude negatively or positively. *J. eBusiness*, 10(1):38–42.

Ho-Dac, N. N., Carson, S. J., and Moore, W. L. (2013). The effects of positive and negative online customer reviews: Do brand strength and category maturity matter? *Journal of Marketing*, 77(6):37–53.

Hu, N., Pavlou, P. A., and Zhang, J. (2017). On self-selection biases in online product reviews. *MIS Quarterly*, 41(2):449–471.

Hu, N., Pavlou, P. A., and Zhang, J. J. (2009). Why do online product reviews have a J-shaped distribution? Overcoming biases in online word-of-mouth communication. *Communications of the ACM*, 52(10):144–147.

Huang, N., Hong, Y., and Burtch, G. (2017). Social network integration and user content generation: Evidence from natural experiments. *MIS Quarterly*, 41(4):1035–1058.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292.

Kosslyn, S. M. (1994). *Elements of Graph Design*. New York: W.H. Freeman.

Krider, R. E., Raghubir, P., and Krishna, A. (2001). Pizzas: π or square? psychophysical biases in area comparisons. *Marketing Science*, 20(4):405–425.

Kronrod, A. and Danziger, S. (2013). "Wii will rock you!" the use and effect of figurative language in consumer reviews of hedonic and utilitarian consumption. *Journal of Consumer Research*, 40(4):726–739.

Kuksov, D. and Xie, Y. (2010). Pricing, frills, and customer ratings. *Marketing Science*, 29(5):925–943.

Lem, S., Onghena, P., Verschaffel, L., and Van Dooren, W. (2014). Interpreting histograms. as easy as it seems? *European journal of psychology of education*, 29(4):557–575.

Li, X. and Hitt, L. M. (2008). Self-selection and information role of online product reviews. *Information Systems Research*, 19(4):456–474.

Liu, Y. (2006). Word of Mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing*, 70(3):74–89.

Luca, M. and Smith, J. (2013). Salience in quality disclosure: Evidence from the US news college rankings. *Journal of Economics & Management Strategy*, 22(1):58–77.

Luca, M. and Zervas, G. (2016). Fake it till you make it: Reputation, competition, and yelp review fraud. *Management Science*, 62(12):3412–3427.

Lurie, N. H. and Mason, C. H. (2007). Visual representation: Implications for decision making. *Journal of Marketing*, 71(1):160–177.

Mayzlin, D., Dover, Y., and Chevalier, J. (2014). Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review*, 104(8):2421–55.

Moe, W. W. and Trusov, M. (2011). The value of social dynamics in online product ratings forums. *Journal of Marketing Research*, 48(3):444–456.

Moore, S. G. (2015). Attitude predictability and helpfulness in online reviews: The role of explained actions and reactions. *Journal of Consumer Research*, 42(1):30–44.

Moore, S. G. and Lafreniere, K. C. (2020). How online word-of-mouth impacts receivers. *Consumer Psychology Review*, 3(1):34–59.

Mudambi, S. M. and Schuff, D. (2010). Research note: What makes a helpful online review? A study of customer reviews on Amazon. com. *MIS Quarterly*, 34(1):185–200.

Newman, G. E. and Scholl, B. J. (2012). Bar graphs depicting averages are perceptually misinterpreted: The within-the-bar bias. *Psychonomic bulletin & review*, 19(4):601–607.

Parkhurst, D., Law, K., and Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, 42(1):107–123.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization*, 3(4):323–343.

Raghubir, P. and Krishna, A. (1999). Vital dimensions in volume perception: Can the eye fool the stomach? *Journal of Marketing Research*, 36(3):313–326.

Rosario, A. B., Sotgiu, F., Valck, K. D., and Bijmolt, T. H. (2016). The effect of electronic Word of Mouth on sales: A meta-analytic review of platform, product, and metric factors. *Journal of Marketing Research*, 53(3):297–318.

Savage, L. J. (1972). *The Foundations of Statistics*. Courier Corporation.

Sen, S. and Lerman, D. (2007). Why are you telling me this? An examination into negative consumer reviews on the web. *Journal of Interactive Marketing*, 21(4):76–94.

Shen, W., Hu, Y. J., and Ulmer, J. R. (2015). Competing for attention: An empirical study of online reviewers' strategic behavior. *MIS Quarterly*, 39(3):683–696.

Sherman, S. J., Cialdini, R. B., Schwartzman, D. F., and Reynolds, K. D. (1985). Imagining can heighten or lower the perceived likelihood of contracting a disease: The mediating effect of ease of imagery. *Personality and Social Psychology Bulletin*, 11(1):118–127.

Simonson, I. (2016). Imperfect progress: An objective quality assessment of the role of user reviews in consumer decision making, a commentary on de Langhe, Fernbach, and Lichtenstein. *Journal of Consumer Research*, 42(6):840–845.

Spence, I. (1990). Visual psychophysics of simple graphical elements. *Journal of experimental psychology: Human perception and performance*, 16(4):683.

Stone, E. R., Yates, J. F., and Parker, A. M. (1997). Effects of numerical and graphical displays on professed risk-taking behavior. *Journal of Experimental Psychology: Applied*, 3(4):243.

Sun, M. (2012). How does the variance of product ratings matter? *Management Science*, 58(4):696–707.

Torralba, A., Oliva, A., Castelhano, M. S., and Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, 113(4):766.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323.

Tversky, B. (1997). Cognitive principles of graphic displays. In *Paper presented at the Association for the Advancement of Artificial Intelligence Workshop on Diagrammatic Reasoning*. Cambridge, MA.

von Neumann, J., Morgenstern, O., and Kuhn, H. W. (2007). *Theory of Games and Economic Behavior*. Princeton University Press.

Wang, C. A., Zhang, X. M., and Hann, I.-H. (2018). Socially nudged: A quasi-experimental study of friends' social influence in online product ratings. *Information Systems Research*, 29(3):641–655.

Xiong, G. and Bharadwaj, S. (2014). Prerelease buzz evolution patterns and new product performance. *Marketing Science*, 33(3):401–421.

Ye, S., Gao, G., and Viswanathan, S. (2014). Strategic behavior in online reputation systems. *Mis Quarterly*, 38(4):1033–1056.

Yin, D., Bond, S. D., and Zhang, H. (2014). Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. *MIS quarterly*, 38(2):539–560.

Yin, D., Mitra, S., and Zhang, H. (2016). When do consumers value positive vs. negative reviews? an empirical investigation of confirmation bias in online word of mouth. *Information Systems Research*, 27(1):131–144.

Zacks, J. and Tversky, B. (1999). Bars and lines: A study of graphic communication. *Memory & cognition*, 27(6):1073–1079.

Zhu, F. and Zhang, X. M. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing*, 74(2):133–148.

Zhu, L., Yin, G., and He, W. (2014). Is this opinion leader's review useful? Peripheral cues for online review helpfulness. *Journal of Electronic Commerce Research*, 15(4):267.

Appendix A: The Residual Term in the Taylor Expansion

In the Taylor expansion of the transformation function of probabilities

$$t(p_i) = t(0) + t'(0)p_i + \frac{1}{2}t''(0)p_i^2 + R_2(p_i) \tag{5}$$

The reminder can be represented as:

$$R_2(p_i) = \frac{1}{6}t'''(\xi_i)p_i^3, \qquad \epsilon_i \in [0, p_i] \tag{6}$$

The subjective weighting is

$$w(p_i|p_{-i}) = \frac{t'(0)p_i + \frac{1}{2}t''(0)p_i^2 + \frac{1}{6}t'''(\xi_i)p_i^3}{t'(0) + \frac{1}{2}t''(0)\sum_j p_j^2 + \sum_j \frac{1}{6}t'''(\xi_j)p_j^3} \tag{7}$$

$$w(p_i|p_{-i}) = \frac{p_i + \lambda p_i^2 + \lambda\theta(\xi_i)p_i^3}{1 + \lambda\sum_j p_j^2 + \sum_j \lambda\theta(\xi_j)p_j^3} \tag{8}$$

Where,

$$\lambda = \frac{1}{2}\frac{t''(0)}{t'(0)} \tag{9}$$

$$\theta(\xi_i) = \frac{1}{3}\frac{t'''(\xi_i)}{t''(0)} \tag{10}$$

The error caused by the reminder is

$$\epsilon(p_i|p_{-i}) = w(p_i|p_{-i}) - \frac{p_i + \lambda p_i^2}{1 + \lambda\sum_j p_j^2} \tag{11}$$

$$\epsilon(p_i|p_{-i}) = \frac{\lambda\theta(\xi_i)p_i^3(1 + \lambda\sum_j p_j^2) - (p_i + \lambda p_i^2)\sum_j \lambda\theta(\xi_j)p_j^3}{(1 + \lambda\sum_j p_j^2)(1 + \lambda\sum_j p_j^2 + \sum_j \lambda\theta(p_i)p_i^3)} = \frac{A}{B} \tag{12}$$

The rating difference due to the error of the reminders can be expressed as

$$\Delta\mu = \sum_i \epsilon(p_i|p_{-i})x_i = \frac{1}{B}(\sum_i \lambda\theta(\xi_i)p_i^3 x_i \; - \sum_i \lambda\theta(\xi_i)p_i^3 \sum_j p_j x_j + \lambda^2 \sum_i \theta(\xi_i)p_i^3 x_i \sum_j p_j^2 - \lambda^2 \sum_i \theta(\xi_i)p_i^3 \sum_j p_j^2 x_j) \tag{13}$$

$$\Delta\mu = \frac{1}{B}(\sum_i \lambda\theta(\xi_i)p_i^3(x_i - \overline{x}) + \sum_i \lambda^2\theta(\xi_i)p_i^3(x_i \sum_j p_j^2 - \sum_j p_j^2 x_j)) \tag{14}$$

$$\Delta\mu = \frac{1}{B}(\sum_i \lambda\theta(\xi_i)p_i^3(x_i - \overline{x}) + \sum_i \lambda^2\theta(\xi_i)p_i^3(x_i \sum_j p_j^2 - \sum_j p_j^2 x_j + \sum_j p_j^2(x_j - \overline{x}) - \sum_j p_j^2(x_j - \overline{x}))) \tag{15}$$

$$\Delta\mu = \frac{1}{B}(1 + \lambda \sum_j p_j^2) \sum_i \lambda\theta(\xi_i)p_i^3(x_i - \overline{x}) - \frac{1}{B} \sum_j p_j^2(x_j - \overline{x}) \sum_i \lambda^2\theta(\xi_i)p_i^3 \tag{16}$$

First of all, if there is no systematic bias away from the second order expansion,

$$\sum_k \theta_k(\xi_{ik}) = 0. \tag{17}$$

The value k here refers to different decision makers. We can see that there will be no systematic bias caused by the reminder error.

$$\sum_k \Delta\mu_k = 0 \tag{18}$$

If there is some systematic bias on the third derivative of the transformation function, the utility function can be expressed as:

$$\Delta\mu = (k_1 - k_2)\mathrm{VDB} \tag{19}$$

where,

$$k_1 = \frac{\sum_i \theta(\xi_i) p_i^3 (x_i - \overline{x})}{\sum_j p_i^2 (x_i - \overline{x})} \frac{1 + \lambda \sum_i p_i^2}{1 + \lambda \sum p_i^2 + \lambda \sum_i \theta(\xi_i) p_i^3} \tag{20}$$

$$k_2 = \frac{\lambda \sum_i \theta(\xi_i) p_i^3}{1 + \lambda \sum p_i^2 + \lambda \sum_i \theta(\xi_i) p_i^3} \tag{21}$$

When the curve of the transformation function does not deviate far from our assumption:

$$\theta(\xi_i) \ll 1 \tag{22}$$

$$k_1 \ll 1, \quad k_2 \ll 1 \tag{23}$$

$$|\Delta\mu| \ll |\mathrm{VDB}| \tag{24}$$

Appendix B: Comparison of Decision Bias Theories

In the following table, we show that our model is different from previous theories about decision biases, including von Neumann et al. (2007), Savage (1972), Kahneman and Tversky (1979), Birnbaum and Stegner (1979), Quiggin (1982), Tversky and Kahneman (1992), and Fisher et al. (2018).

Appendix C: Additional Experiment Details

Stimuli for Study 6

	(1) EU	(2) SEU	(3) PT	(4) CWT	(5) RD / CPT	(6) BB	(7) VDB
Subjective Probableities	No	Yes	Yes	Yes	Yes	Yes	Yes
Number of Events	Multiple	Multiple	Two	Two	Multiple	Multiple	Multiple
Weightings Sum to 1	Yes	Yes	No	Yes	Yes	Yes	Yes
Weightings Distorted by Outcomes' Utilities	No	No	No	Yes	Yes	No	No
Weightings Distorted by Other Probability Components	No	No	No	No	No	No	Yes
Precise Mathematical Measure	Yes	No	No	No	No	No	Yes

(1) EU	Expected Utility Theory (Von Neumann & Morgentern 1947)
(2) SEU	Subjective Expected Utility Theory (Leonard Savage 1954)
(3) PT	Prospect Theory (Kahneman & Tversky 1979)
(4) CWT	Configural Weighting Theory (Birnbaum & Stegner 1979)
(5) RD / CPT	Rank Dependent (Quiggin 1982) / Cumulative Prospect Theory (Kahneman & Tversky 1992)
(6) BB	Binary Bias (Fisher, Newman, Dhar & van Osselaer 2018)
(7) VDB	Visual Distortion Bias

Table 2: Comparison of Studies of Objective and Subjective Expected Utility

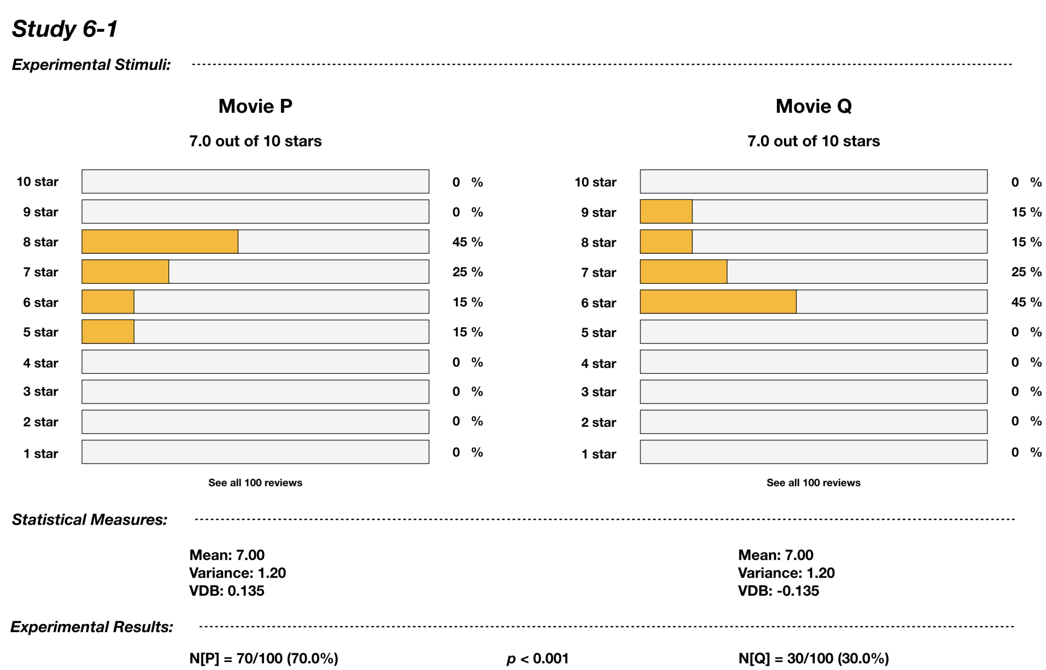

Figure 18: Stimuli for Study 6-1

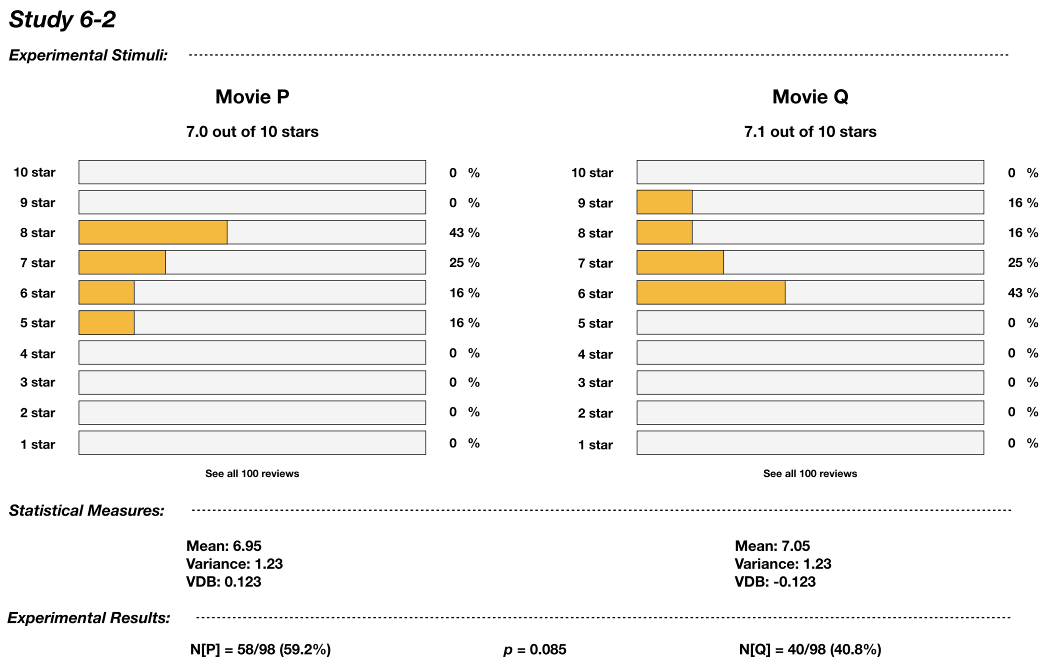

Figure 19: Stimuli for Study 6-2

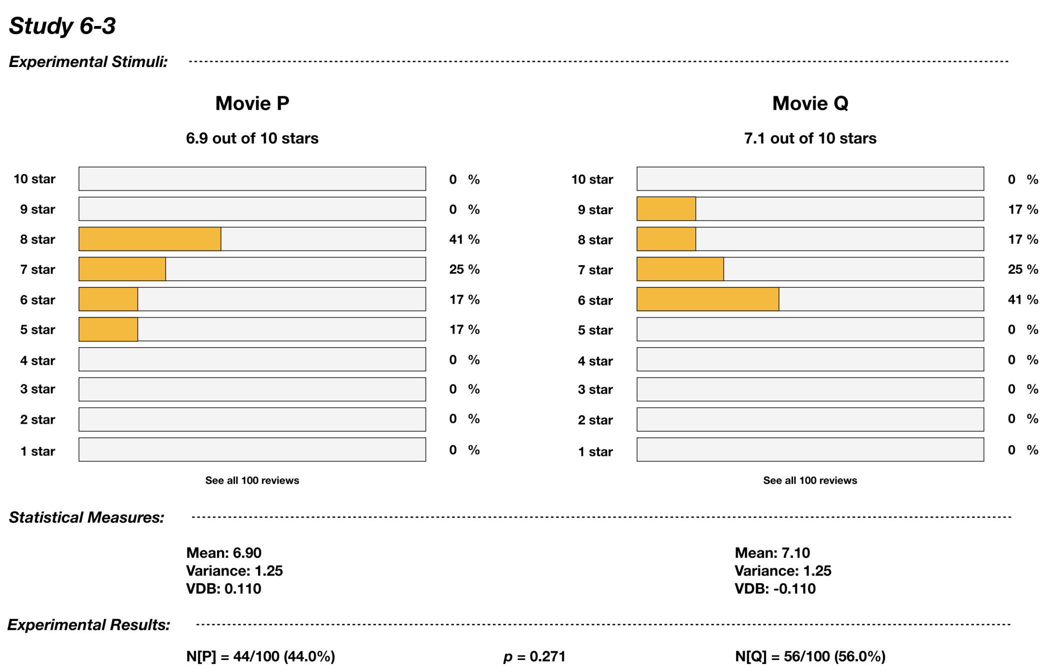

Figure 20: Stimuli for Study 6-3

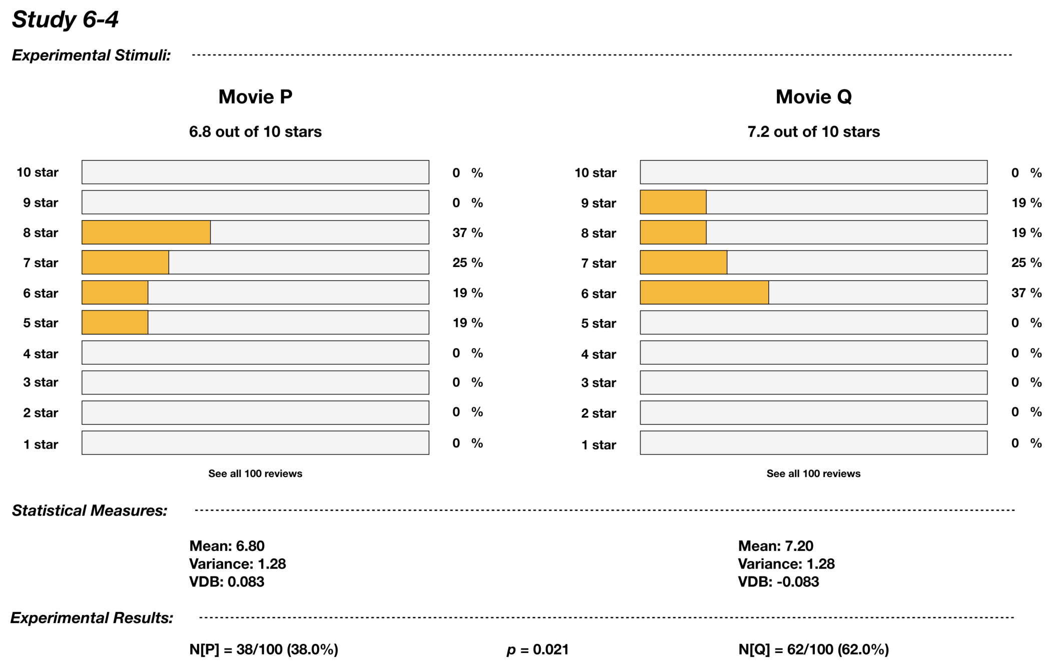

Figure 21: Stimuli for Study 6-4

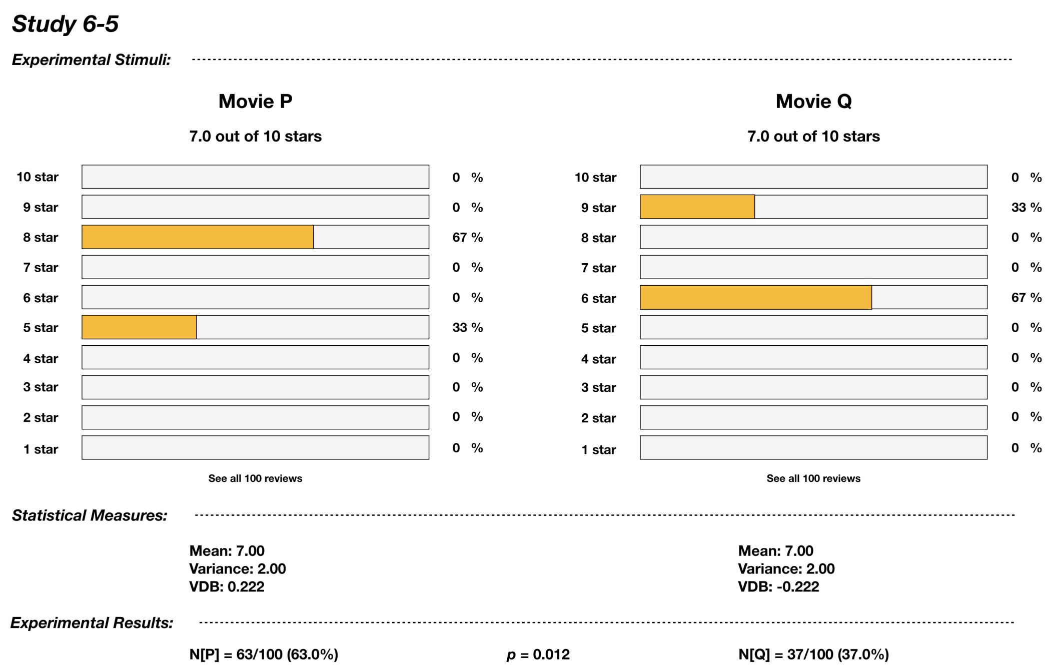

Figure 22: Stimuli for Study 6-5

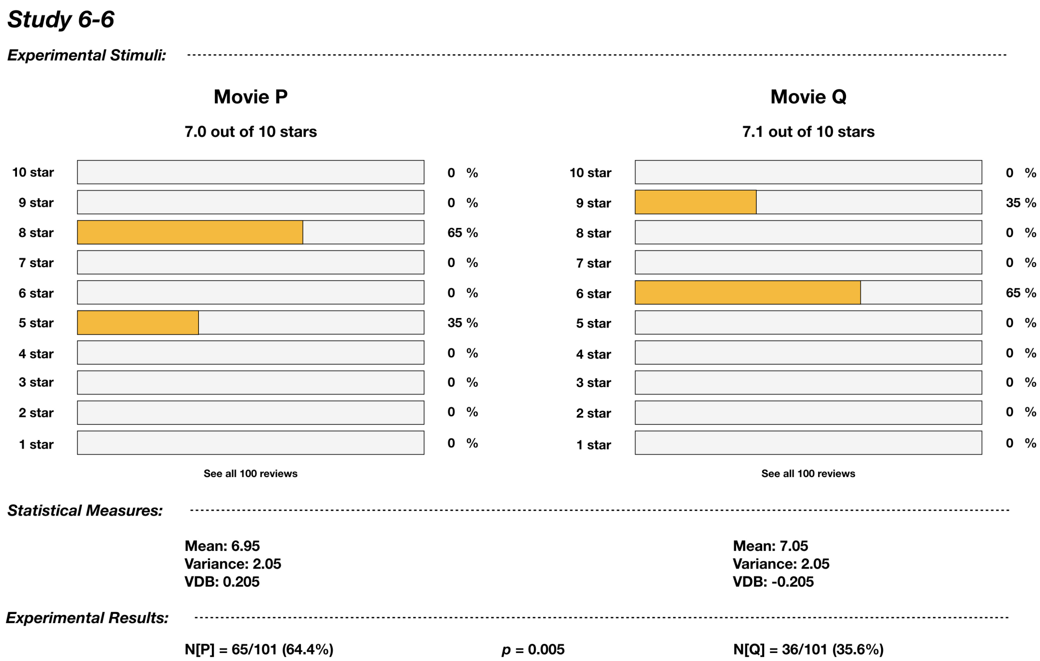

Figure 23: Stimuli for Study 6-6

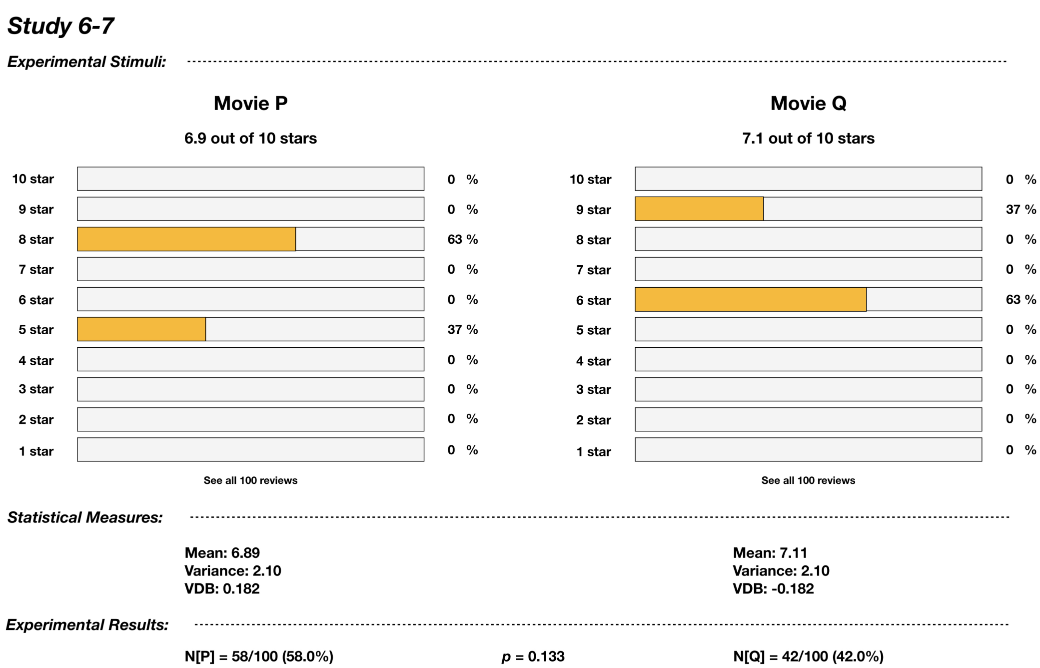

Figure 24: Stimuli for Study 6-7

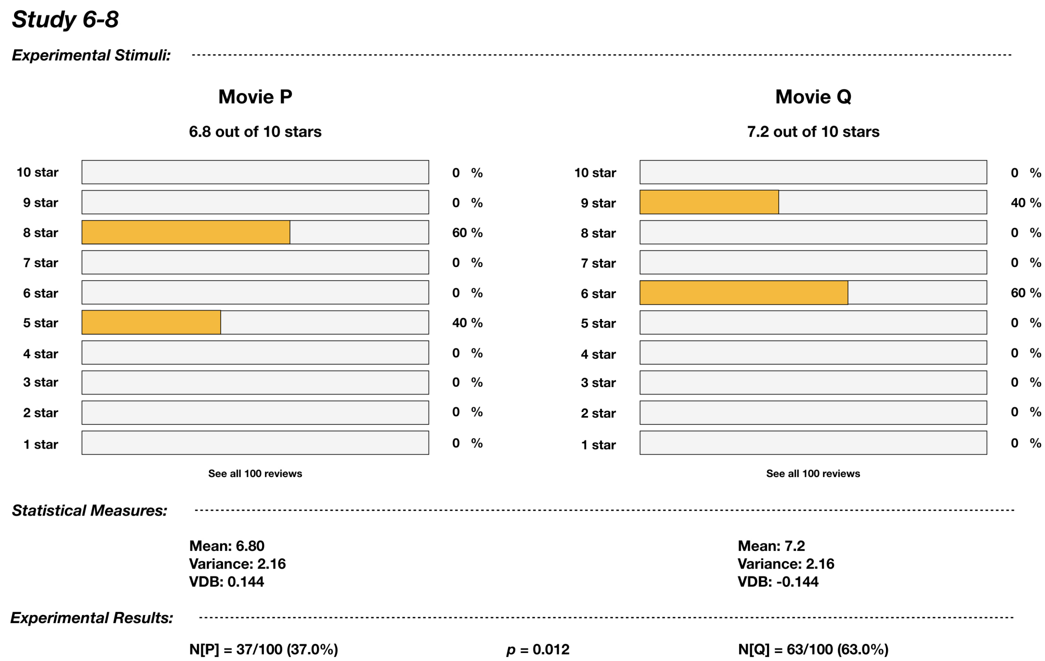

Figure 25: Stimuli for Study 6-8

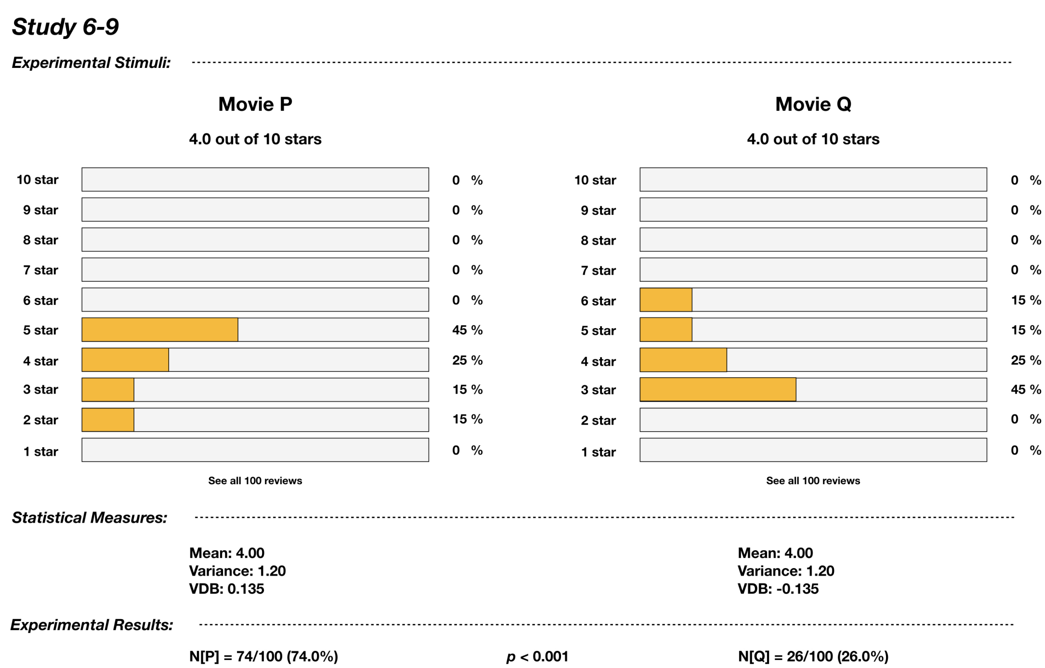

Figure 26: Stimuli for Study 6-9

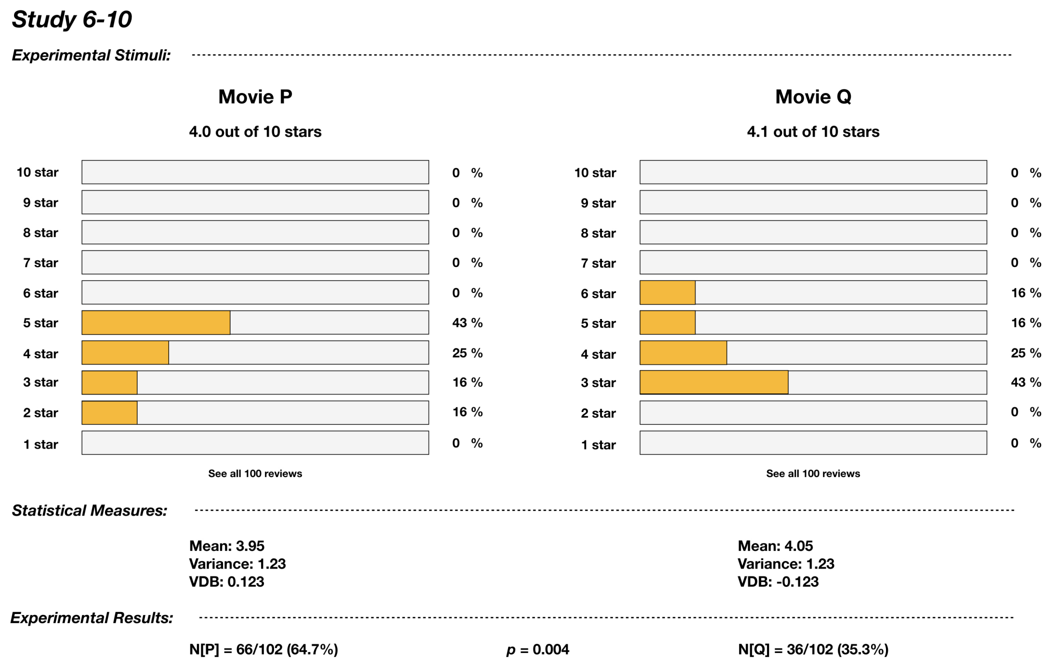

Figure 27: Stimuli for Study 6-10

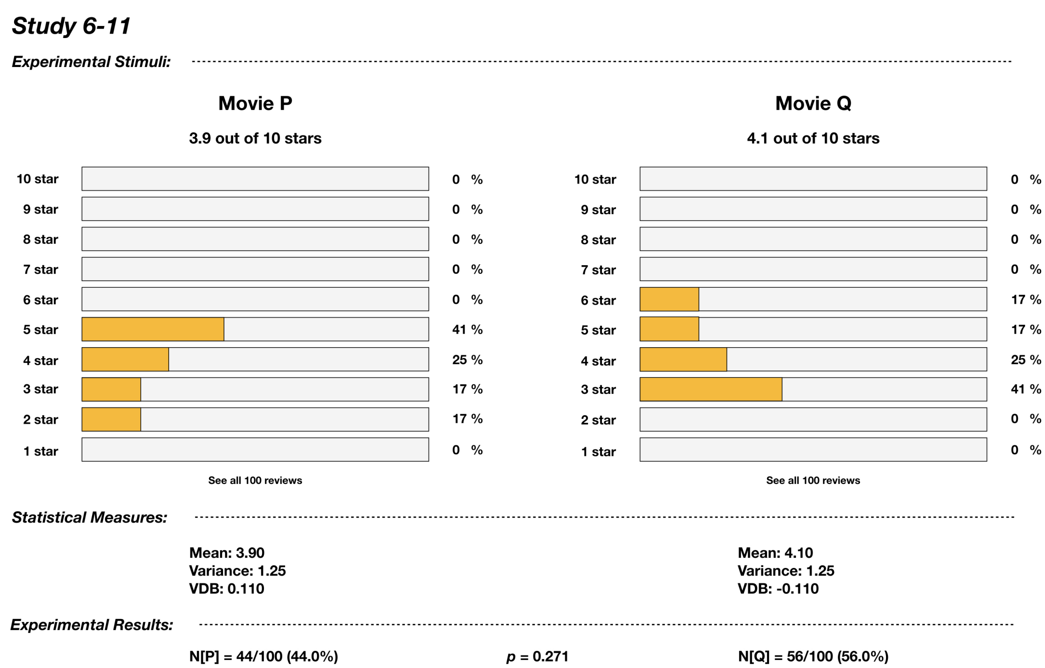

Figure 28: Stimuli for Study 6-11

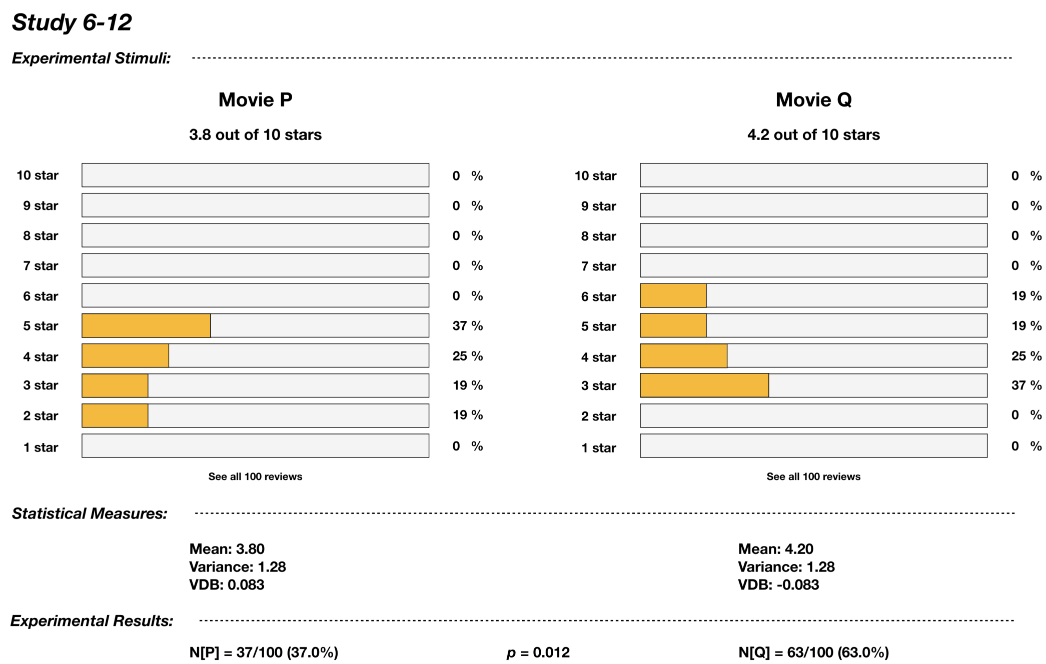

Figure 29: Stimuli for Study 6-12

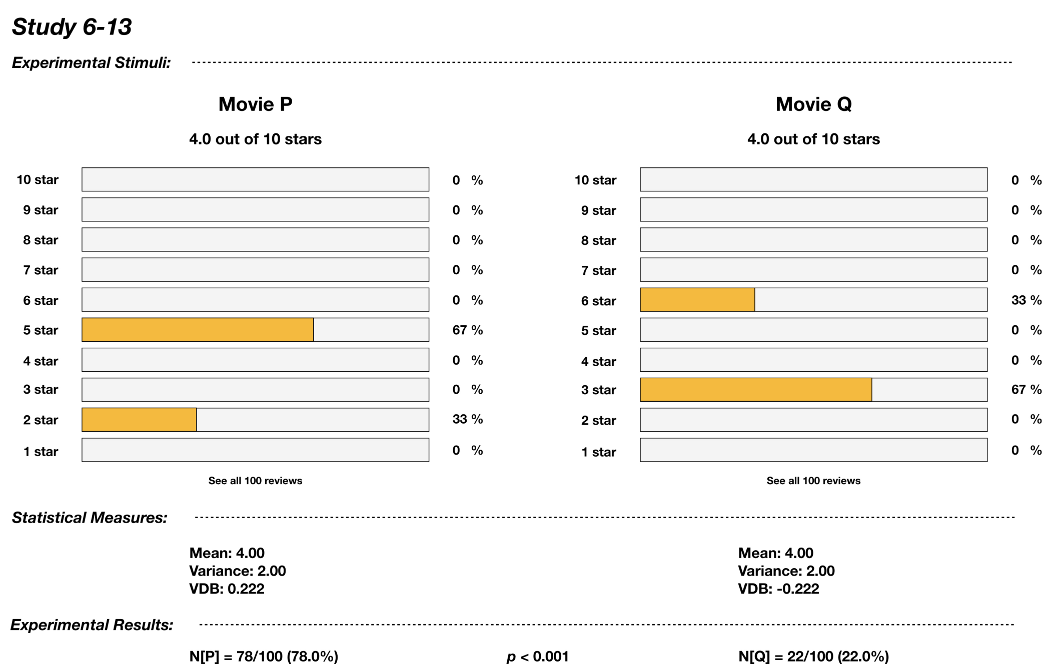

Figure 30: Stimuli for Study 6-13

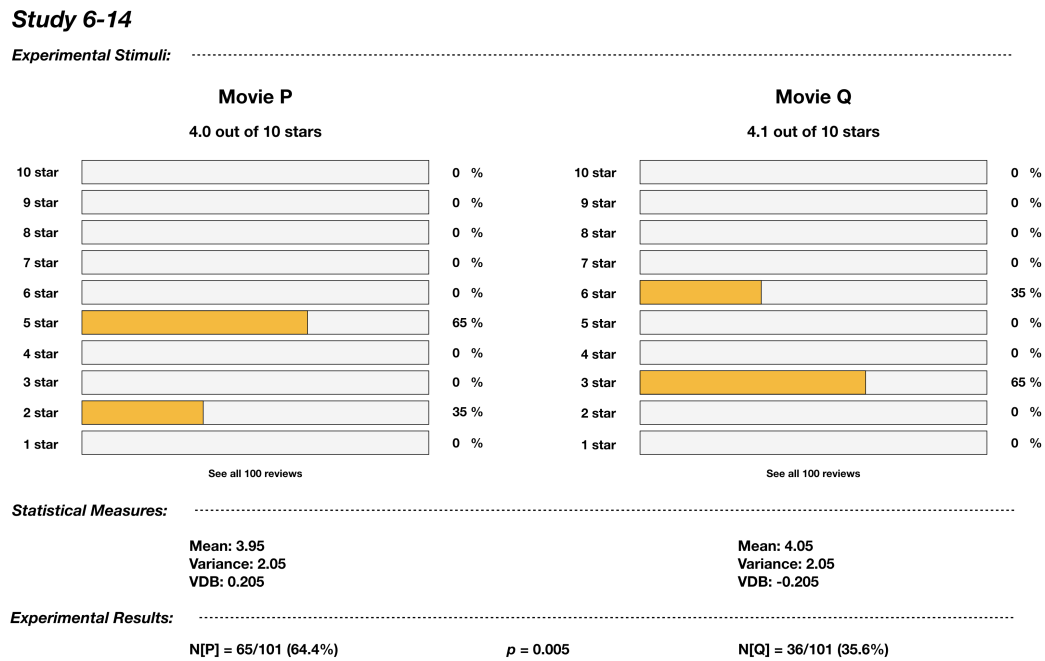

Figure 31: Stimuli for Study 6-14

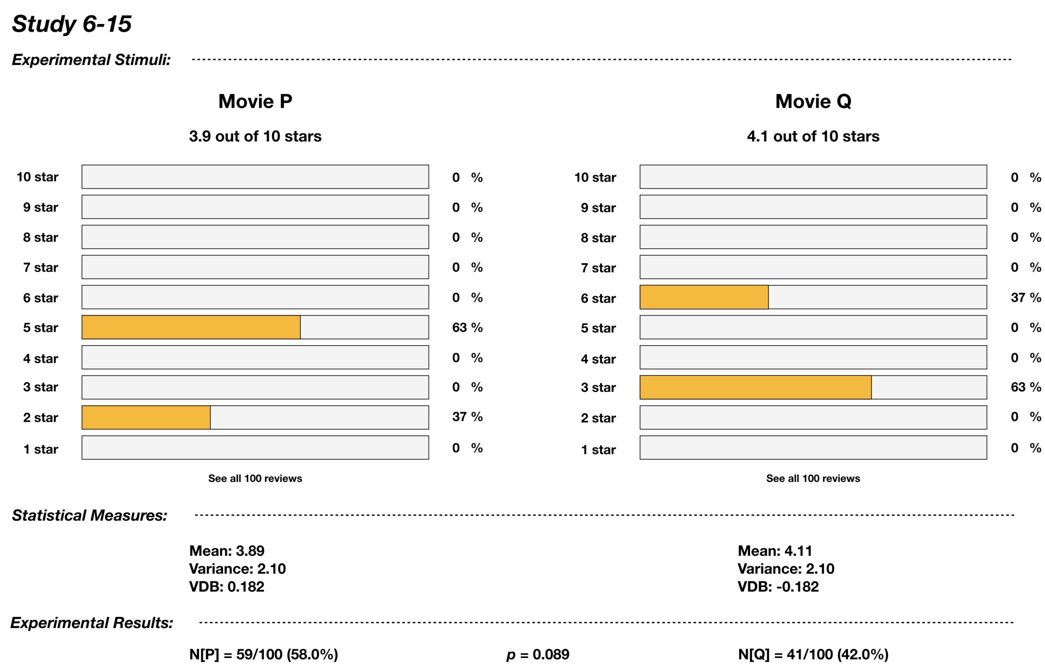

Figure 32: Stimuli for Study 6-15

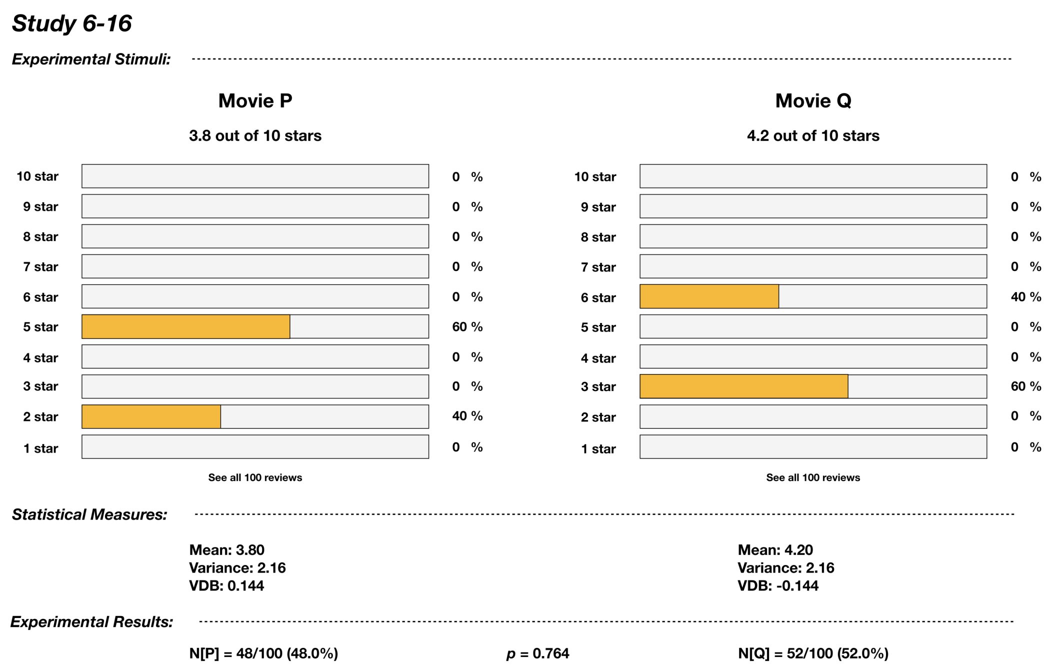

Figure 33: Stimuli for Study 6-16