Structural Estimation of Expert Strategic Bias: the Case of Movie Reviewers*

Fanny Camara[†] and Nicolas Dupuis[‡]

February 17, 2014

Abstract

We develop the first structural estimation of reputational cheap-talk games using data on movie reviews released in the US between 2004 and 2013. Our approach allows us to jointly identify and estimate movies' priors, as well as movie reviewers' abilities and strategic biases. We find that reviewers adopt reporting strategies that are consistent with the predictions of the literature on reputational cheap-talk. The average conservatism bias for low prior movies lies between 8 and 11%, depending on the specifications of the model. The average conservatism bias for high prior movies ranges from 13 to 15%. Moreover, we find a significant, albeit small, effect of the reputation of the reviewers on their strategies, indicating that incentives to manipulate demand in order to prevent reputation updating are present in this industry. Our estimation takes into account and quantifies potential conflicts of interest that might arise when the movie reviewer belongs to the same media outlet as the film under review. Out-of-sample predictions show that our model performs better than alternative specifications without reputational concerns.

Keywords: Structural estimation, Reputational cheap-talk game, Delegated expertise, Film Industry.

JEL classification: C21, L15, L82, Z11

*We would like to thank Peter Arcidiacono, Christian Bontemps, Stefano DellaVigna, Pierre Dubois, Guido Friebel, Yinghua He, Marc Ivaldi, Bruno Jullien, Margaret Kyle, Thierry Magnac, Mar Reguant, François Salanié, Florian Schuett, Sergio Urzua, and audiences at several conferences and seminars for their precious advice. We would also like to thank Julien Arramon-Tucoo for his valuable help with the data. All remaining errors are our own.

[†]Ph.D. Candidate, Toulouse School of Economics, 21 alle de Brienne, 31015 Toulouse Cedex 6, France. *e-mail: fanny.camara@tse-fr.eu*

[‡]Ph.D. Candidate, Toulouse School of Economics, 21 alle de Brienne, 31015 Toulouse Cedex 6, France. *e-mail: nicolas.dupuis@tse-fr.eu*

1 Introduction

On that point, Socrates, I have heard that one who is to be an orator does not need to know what is really just, but what would seem just to the multitude who are to pass judgment, and not what is really good or noble, but what will seem to be so; for they say that persuasion comes from what seems to be true, not from the truth.

– Plato, *Phaedrus*, 260a

We use a reputational cheap talk framework to model the reviewing behavior of movie critics. We develop a new structural approach to quantify the strategic biases of movie reviewers which arise from their reputational concerns. Our approach allows us to separately identify and estimate individual reviewers' abilities and biases. Comparing the predicting power of our model to the one of models without reputational concerns, we find empirical evidence that career concerns do shape the behavior of movie reviewers.

Cheap-talk games apply to economic situations in which an expert with private information about the state of the world sends a prediction about this state to some receivers. The expert sends the prediction at no cost and cannot base this report on any verifiable (hard) evidence. In this paper, we focus on the subclass of reputational cheap-talk games in which the expert seeks to maximize her reputation for accuracy. The expert's reputation corresponds to the receivers' belief about the precision of her private information.

Movie reviewers watch films before their official release and get a private signal on their quality whose accuracy depends on their ability. The career of these reviewers is built on their reputation for accuracy. Any expert assessing the quality of experience goods in the absence of hard evidence on quality is subject to the same incentives. Financial analysts sell their advice on investment opportunities about which they are supposed to have some private information. A higher reputation for good advice allows them to charge higher prices. Academic referees in peer-reviewed scientific journals give their opinion on whether or not

a paper is publishable based on their personal assessment of its scientific quality. Signaling their good judgement to the editor can help their academic career.

In their seminal paper on reputational cheap-talk, Ottaviani and Sørensen (2006) show that experts may disregard noisy signals and conform to the prevalent opinion in order to pass for good predictors of the state of the world. Hence, when there is a strong prior belief on the state of the world being high and the expert receives a private signal supporting the opposite, she has the incentive to lie and to pretend that she received a high signal. This tendency for the expert to stick to extreme priors is referred to as conservatism bias. In the most informative equilibrium, despite this incentive, the high ability expert truthfully reveals her signal whereas the unskilled one conforms to the common prior with some positive probability. In this binary state setup, when the public belief about the state is more balanced, both types of experts are inclined to truthfully reveal their signals.

Camara and Dupuis (2013) and Mariano (2012) study a similar game in which the state of the world is imprecisely revealed. These papers endogeneize the precision with which the receivers can observe the state after its realization. In particular, the expert can make it harder for the receivers to observe the true state by convincing them that it is low. For instance, movie reviewers can discourage consumers to see a movie by writing very bad reviews. It is even more true if the movie is obscure to begin with. In this case, the scarce audience of the movie prevents the market from learning about its quality. Advice not to invest lead to no investment and thus to the failure of the project, regardless of its quality, which then becomes unobservable. When unsure about the accuracy of her high signal, the expert has the incentive to send a low report in order to garble her reputation update. We call manipulation bias this extra incentive for experts to over-report the low state. This manipulation bias is stronger for influential experts, i.e. experts who enjoy a high reputation, as their recommendations are more likely to change drastically receivers' beliefs. Simultaneous competition among experts has no effect on conservatism bias but can

mitigate the manipulation bias as the extent to which each expert's recommendation affects beliefs diminishes with the presence of contradictory recommendations. Hence, contrary to the case of perfect ex-post revelation of the state of the world, simultaneous competition, reputation, or the polarity of the prior (i.e. whether it is low or high) do have an impact on the experts' strategies.

In this paper, we develop the first structural estimation of cheap-talk games. We bring the theoretical model developed in Camara and Dupuis (2013) to data on movie reviews. We test whether experts misreport their private information and we show that the reporting strategies are in line with the predictions of the literature on reputational cheap-talk. On top of conservatism bias, our estimation shows that reviewers are also subject to a manipulation bias. We quantify the extent of misreporting for each expert. We recover the unobserved ability of each reviewer in our sample. Moreover, our estimation takes into account and quantifies potential conflicts of interest that might arise when the movie reviewer belongs to the same media outlet as the film under review.

Studying strategic biases in cheap-talk games is of great importance because they can lead to welfare loss. Considering that new entrants in a market usually suffer from a lower prior than incumbents, expert bias can actually lead to additional barriers to entry. Due to strategic biases, the young director with no experience is likely to get harsher reviews than her older colleagues. Properly quantifying these strategic biases is key to acquiring a deep understanding of the role of experts in these markets and preventing practices resulting in welfare losses.

Following Camara and Dupuis (2013), we extend the model to a continuum of expert's abilities and provide a simple characterization of the incentives to be truthtelling. This allows us to characterize a unique set of priors over which every expert, regardless of her ability, is truthtelling in the most informative equilibrium. This truthtelling set is at the core of our estimation strategy.

In our empirical analysis, we control for horizontal differentiation by using data on MPAA ratings that are highly correlated with genres. The quality is the vertical element of differentiation between movies: within each genre, a high quality movie is more likely to please a consumer than a low quality movie. We jointly identify priors, biases and abilities. For given biases and abilities, higher prior movies tend to get more unanimously good reviews. Conversely, once the prior is identified, we retrieve both abilities and biases using a partition on the prior space. This partition is derived from the theoretical model and consists in priors falling within or outside a truthful revelation set. This set is the subset of the prior space on which experts of any ability are truthtelling in equilibrium. Its position in the prior space also depends on the reputation of the expert. Inside the truthful revelation set, absent of bias, the distribution of observed recommendations is determined by priors and experts' abilities. We can therefore recover abilities for given priors. Outside the truthful revelation set, for given priors and abilities, experts' biases are directly identified from the observed recommendations. We observe variation over time in the reputation of each reviewer which shifts their truthful revelation set and then allow us to identify its effect on each reviewer's bias. Our estimation strategy directly proceeds from the identification since we use the same partition structure over the set of priors in a maximum likelihood approach.

We collect a nearly exhaustive data set on movies released in the US between 1990 and 2013 that contains the main information on directors, production companies, budgets, exact release dates, and genres. For each of those movies we gather the reviews published on *rottentomatoes.com*, the reference website displaying critics from the most influential movie reviewers. The reviews posted on this website have a binary structure (the movie is either *fresh* or *rotten*) which perfectly matches our theoretical model. Along with the name of the reviewer and her positive or negative recommendation, we collect the medium in which the review was published and the date of the publication. We use *Google trend* to retrieve monthly data on the number of Google searches which measure reviewers' reputation.

We find movie reviewers' abilities ranging from 62% to 90%, meaning that the least able movie reviewer in our sample receives the correct private signal 6 times out of 10 whereas the most able one is correct 9 times out of 10. The average conservatism bias for low prior movies, i.e. the average probability that the experts in our sample transform a positive signal into a bad review when the prior is low is between 8 and 11%, depending on the specifications of the model. The average conservatism bias for high prior movies ranges from 13 to 15%. For some reviewers in our sample this conservatism bias goes up to 40%. Moreover, we find a significant, albeit small, effect of the reputation of reviewers on their strategies, indicating that incentives to manipulate demand in order to prevent reputation updating are present in this industry. We estimate an average probability of giving a good review despite a negative signal of 5% among reviewers due to conflicts of interest.

Our estimation strategy relies on weak assumptions: we assume that, for a given prior, if an equilibrium exists in which experts of all possible abilities are truthtelling, then this equilibrium is played. An alternative for the movie reviewers would be to play a babbling equilibrium. Since we observe that movie reviewers send reviews correlated to the true quality[1], we think this is highly unlikely. Moreover, as shown by other papers in the literature (Basuroy et al. (2003), Reinstein and Snyder (2005), Basuroy and Ravid (2013)), consumers tend to take into account movie reviews when they make their purchasing decision. This behavior is inconsistent with a babbling equilibrium.

Our estimation of the bias does not rely on any additional assumptions concerning which equilibrium is played outside the truthtelling set, which is fortunate considering that cheap-talk games usually feature multiple equilibria.

A reduced-form approach would have required to accurately estimate the quality of each movie to lift any endogeneity bias. By contrast, this structural estimation avoids making

[1]A previous version of this paper, featuring a reduced-form estimation of expert bias, provides some evidence of this phenomenon.

any assumptions on the quality of movies. Moreover, a reduced-form approach would have only allowed us to estimate an aggregate bias of movie reviewers, whereas our method allows us to identify and estimate both individual abilities and strategic bias.

In order to provide evidence of the performance of our econometric model, we carry out an estimation of our model on a random subsample of our data set and conduct out-of-sample predictions. We find that the predicted distribution of reviews matches closely the one observed in the data. We also compare our model to two alternative specifications: one ruling out strategic biases and one assuming that all experts have the same ability and this ability is common knowledge. The first one corresponds to a story in which experts differ in their ability but are nonetheless truthful when sending their recommendations. The second one corresponds to a story in which experts are truthful because their payoff does not depend on how well informed they are. We find that our model performs better than these two in predicting the outcomes in the out-of-sample data, suggesting that strategic biases and career concerns do play an important role in movie reviewing.

Technically, our estimation method follows the literature on the estimation of voting games by building on the technique developed by Iaryczower and Shum (2012b). They estimate the political bias of supreme court judges who aim to take the right decision between rejecting or confirming the decision of the lower court. The main conceptual difference between our two approaches is that political biases are exogenous whereas our strategic bias is endogenous as it depends on the prior as well as the ability and reputation of the expert. They rely on a two-step approach in which biases and abilities are recovered from a first-step estimation of conditional voting probabities. By contrast, our identification and estimation strategies rely on a partition of the set of priors in which we directly embed the structure of the theoretical model.

Assessing the importance of strategic bias in expert reviews is critical as they may have a non negligible role in the failure or success of new products, as shown in Reinstein and

Snyder (2005) or in Boatwright et al. (2007). Gentzkow and Shapiro (2006) show that media outlets tend to conform their news coverage to their readership's taste in order to build their reputation as reliable news sources. Deviation from truthful revelation decreases with the verifiability of the state of the world and with competition. Their framework differs from ours in that the expert cannot influence the precision with which the state of the world is observed. Some empirical papers have tackled other potential sources of bias: DellaVigna and Kennedy (2011) or Dobrescu et al. (2012) have studied the impact of media concentration on reviews respectively in the motion pictures and books industries. Both papers find some evidence that reviewers tend to give better reviews to products from the same media outlet, although the latter attributes this effect to similarities in tastes. The latter also finds some evidence that professional reviewers are less favorable to first time authors compared to consumer reviews. However, they only recover an aggregate effect, do not attribute this to strategic bias and do not control for the possibility that professional and amateur reviewers grade differently. All those papers provide atheoretical statistical evidence rather than structural estimations and cannot quantify the extent of the biases.

Theoretical predictions and our empirical strategy are respectively presented in sections 2 and 3. We introduce our dataset in section 4 and our empirical analysis in section 5. Out-of-sample predictions are provided in section 6. Section 7 provides a short conclusion.

2 A reputational cheap talk framework

In this section, we present the reputational cheap-talk game that we want to estimate and characterize the truthful revelation set. This set is defined as all priors supporting a truthful revelation equilibrium for experts of all abilities and is at the core of our identification strategy.

We study a reputational cheap-talk game in which an expert receives a noisy private

information (or signal) about the state of the world and can communicate her information to her audience. For the following, we will assume that the state of the world represents the quality of an experience good while the expert's audience consists of the consumers who can potentially buy the product.

The expert and consumers share a common prior belief about the quality of the product. The report of the expert takes the form of a recommendation about the product. The expert can send a recommendation at no cost. The expert cannot certify that her recommendation matches her private information nor can consumers verify that the expert reported her private information truthfully. The precision of the expert's private information depends on her ability. A more able expert receives a more precise signal.

The expert cares only about her reputation, which is the consumers' belief about her ability. Consumers can base their purchasing decisions on the expert's recommendation. After consumption, each purchaser forms her individual opinion about the product's quality. The aggregation of the individual opinions of the purchasers forms an ex post feedback that the entire consumer population use to compute their posterior belief on the quality. The precision of the ex post feedback increases with the number of consumers who decided to buy the product and could therefore form their opinion on its quality. Ultimately, consumers use their posterior belief on the quality to infer the probability which which the recommendation sent by the expert is correct and update the expert reputation accordingly.

In this game, the expert has two types of incentives not to reveal her private information truthfully. (i) When the prior on the product's quality is extreme and the expert receives a signal that contradicts the prior, she anticipates that her signal is likely to be incorrect and has an incentive to lie and pretend having received the signal that is the most likely to be correct. This yields to a *conservatism bias*. (ii) Since consumers are less likely to buy when they expect the product's quality to be bad, by sending a bad recommendation, the expert can decrease the precision of the ex post feedback and prevent consumers from updating

her reputation. Therefore, the expert has an incentive to over report the low signal. This incentive generates a *manipulation bias*.

The incentive for the expert to overreport the low signal to obfuscate the realization of the state of the world is stronger for highly reputed experts whose recommendation can change more drastically the consumers' updated prior on the quality and then influence more their purchasing decisions. Also simultaneous competition between experts mitigates the manipulation bias since a good recommendation written by a competitive expert reduces the deference effect on consumer demand that a bad recommendation can exert.

2.1 Theoretical Foundations of the Structural Estimation

Our structural estimation relies on a few propositions derived in the theoretical model. We now explicit these propositions. More precisely, we extend the theoretical model in Camara and Dupuis (2013) to a continuum of abilities $t \in \left[\underline{t}, \bar{t}\right]$, $\underline{t}, \bar{t} \in \left(\frac{1}{2}, 1\right)$ and derive a simple characterization of the set of priors over which the experts have the incentives to be truthtelling.

2.1.1 Setup

We assume that the quality, θ, is either good or bad. We denote by Θ the quality space, $\Theta = \{\theta_0, \theta_1\}$. The private signal of the expert, s, is either high or low, $S = \{s_0, s_1\}$. The recommendation of the expert, r, is also binary, $R = \{r_0, r_1\}$. We denote by μ the common prior on the quality being high, $\mu = pr(\theta_1)$. The ability t of the expert is defined as the probability of observing a private signal corresponding to the true quality, $p(s_i|\theta_i)$ in which $i = 0, 1$. We assume that t is private information to the expert. Consumers share a common prior belief about the expert's ability $F(\tilde{t}) = P(t \leq \tilde{t})$ defined on $\left[\underline{t}, \bar{t}\right]$, with an associated density function $f(\tilde{t}) = F'(\tilde{t})$. In the remaining of the paper, we use a tilde to denote consumers' beliefs.

The expert derives a utility $u(\tilde{t})$ of being perceived as having an ability $\tilde{t}$. We assume the expert is risk-neutral, therefore u is linear in $\tilde{t}$. The expert is said to be truthtelling if she sends a recommendation r_i after receiving a signal s_i, for $i = 0, 1$. She has the incentive to be truthtelling when:

$$\mathbb{E}_f(u(\tilde{t})|s_i, r_i, t, \tilde{\sigma}^T) \geq \mathbb{E}_f(u(\tilde{t})|s_i, r_{-i}, t, \tilde{\sigma}^T), \qquad \forall i$$

in which σ^T is the truthtelling strategy. Hence, $\tilde{\sigma}^T$ means that consumers believe the expert is truthtelling. In the following, to simplify notations, we sometimes forget to explicitely mention the distribution on which we compute the expected utility, i.e. f. We also replace the pair $\left\{r_i, \tilde{\sigma}^T\right\}$ by $\tilde{s}_i$, because if consumers think the expert is truthtelling and they receive r_i, they think the expert has received s_i (which is what matters to update her reputation).

After receiving the recommendation of the expert, consumers form an updated prior $\nu(\tilde{s}) = p(\theta_1|\tilde{s})$ about the state of the world, in which $\tilde{s}$ is the consumers' belief about the signal received by the expert, given her recommendation and strategy. The precision of the observation of the state of the world $\tau(.) \geq \frac{1}{2}$ is an increasing function of ν. The idea is that a higher belief on the state of the world being high increases the number of purchases, and a high number of purchases means more accurate consumer reviews. Formally, consumers observe an aggregate signal $X \in \{X_0, X_1\}$ at the end of the game and $p(X_i|\theta_i) = \tau(\nu)$. Models with revelation of the state of the world are equivalent to $\tau(\nu) = 1, \quad \forall \nu$.

Figure 1 presents the timing of the game.

2.1.2 Truthful Revelation Set

This leads us to the following characterization of the truthful revelation set. Its proof can be found in appendix A along the proofs of all propositions in this section:

Proposition 1. *Suppose consumers believe experts of all ability are truthtelling. After re-*

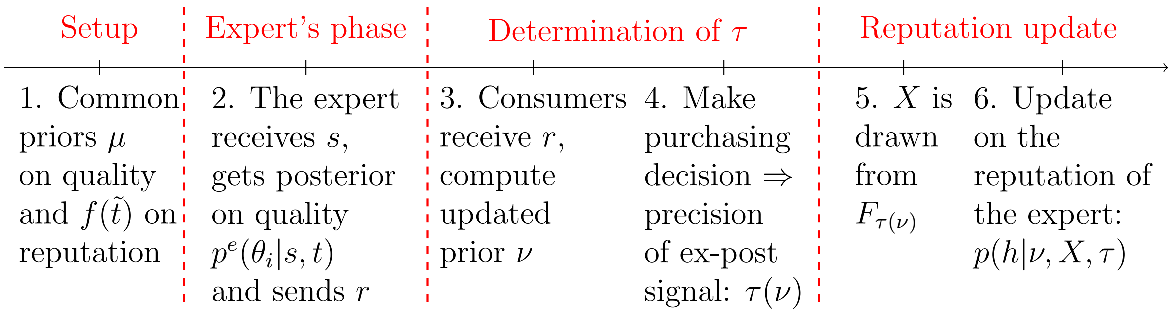

Figure 1: Timing of the game

ceiving $s_i, i \in \{0,1\}$, the expert has no incentives to deviate from truthtelling if and only if:

$$p^e(\tilde{\theta}_i|r_i, s_i, t, \mathbb{E}(\tilde{t})) \geq p^e(\tilde{\theta}_{-i}|r_{-i}, s_i, t, \mathbb{E}(\tilde{t})) \tag{2.1}$$

$\tilde{\theta}_i$ corresponds to the consumers' expectation on the quality. p^e denotes the probability computed by the expert. In words, the expert has no incentives to deviate from truthtelling when she believes that she is more likely to be perceived by consumers as being right when she tells the truth ($r_i = s_i$) than when she lies ($r_i \neq s_i$).

Combining inequalities 2.1 for s_0 and s_1 yields a set of incentive compatible priors depending on the ability of the expert and the consumers' expectation on her ability, which we denote $IC_{t,\mathbb{E}(\tilde{t})}$. As seen in the proof of the proposition 1, only the expected ability of the expert affect the set of incentive compatible priors and not the whole distribution $f(\tilde{t})$.

Corollary 1. *In models with perfect revelation of the state of the world, $IC_{t,\mathbb{E}(\tilde{t})} = [1-t, t]$.*

As shown by corollary 1, when the state of the world is ultimately revealed, the reputation of the expert, $\mathbb{E}(\tilde{t})$, does not affect the set of incentive compatible priors. This set is also centered around $\frac{1}{2}$.

We also show that low ability experts have less incentives to be truthtelling:

Proposition 2. *Fix $\mathbb{E}(\tilde{t})$. Then, $IC_{t_1,\mathbb{E}(\tilde{t})} \subseteq IC_{t_2,\mathbb{E}(\tilde{t})} \Leftrightarrow t_1 \leq t_2$*

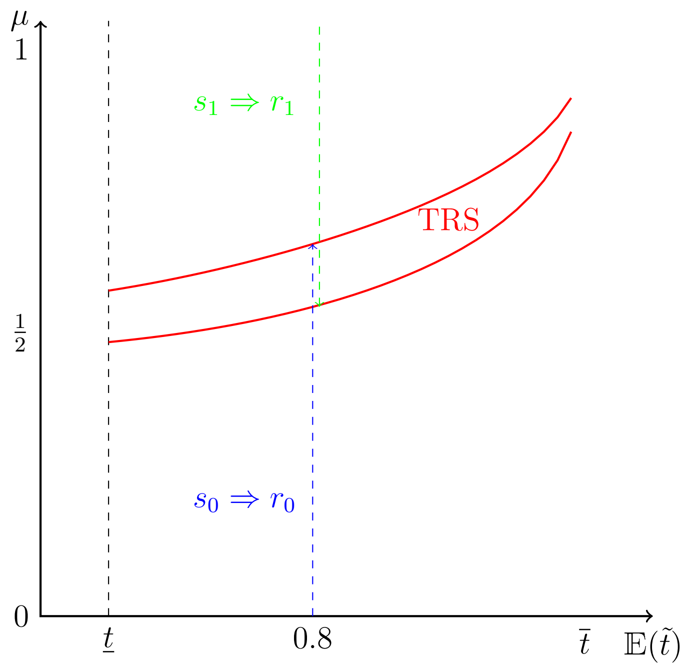

Figure 2: Truthful Revelation Set for $\underline{t} = 0.65$ and $\tau(\nu) = \frac{1+\nu}{2}$. The TRS lies between the two red lines. These lines are determined by the set of incentive compatible priors for s_0, s_1, and all possible values of $\mathbb{E}(\tilde{t})$.

Hence, the set of incentive compatible priors corresponding to the least able expert, i.e. $IC_{\underline{t},\mathbb{E}(\tilde{t})}$, is included in all other sets. It is direct to see that for such priors, truthtelling is an equilibrium for all types of experts. We call this set the *truthful revelation set*.

Figure 2 gives the shape of such a set for different consumer beliefs about the ability of the expert when $\underline{t} = 0.65$.

In this paper, we do not characterize the equilibrium outside the truthful revelation set as it does not affect our estimation strategy.

3 Estimation Strategy

The challenge in estimating biases in the reputational cheap-talk game played by movie reviewers stems from the unobservability of the movies' true quality, the experts' abilities,

and of their private signals. Our estimation strategy has the double advantage not to rely on an estimation of quality to which we would compare the reviews, and to allow us to recover the ability and prior-dependent biases of each expert in our sample. Our estimation consists in simultaneously recovering the prior on the products and the expert-prior-specific conditional probabilities of giving good and bad reviews. The structure of our theoretical model is directly embedded in the estimation process.

3.1 Identification of Ability and Bias

We are able to recover the prior distribution of the quality as the reviews of the movie reviewers, which depend on their private signals, are correlated through the true quality. Indeed, a high quality movie is more likely to generate positive private signals than a low quality one. Holding the biases and abilities constant, the prior is indentified as movies with extreme priors will tend to have more uniform reviews. Once the prior is identified, the partition on the set of priors created by the truthful revelation set allows us to identify each expert's ability and bias. We use observations with priors falling within the truthful revelation set to identify the abilities. For movies whose priors lie in this set, reviewers reveal truthfully their signals and their ability is given by the distribution of their reviews conditional on the quality. Outside the truthful revelation set, the strategic bias is estimated using the difference in the actual distribution of recommendations and the distribution of signals generated by the ability. Finally, the truthtelling set may vary according to the reputation of the expert. We identify the experts' reputation using this exogenous shift in the truthful revelation set, the ability, and the bias.

We now present a more formal argument for the identification. To obtain our result we make the following assumption on the game played by movie reviewers:

Assumption 1. *Experts play truthfully in the truthful revelation set.*

This assumption amounts to say that experts do not play a babbling equilibrium when they review movies whose priors fall within the truthful revelation set. Some empirical evidence support this assumption: Reinstein and Snyder (2005) show that reviews have a significant impact on movie success. If a babbling equilibrium was played, consumers would not make their purchasing decisions according to the reviews. There are also some empirical evidence[2] that movie reviewers on average send informative reviews.

Let us introduce some notations:

- $r_{i,j} \in \{0, 1\}$ is the recommendation or review given by the movie reviewer i to the movie j in our database. It corresponds to the recommendation in the theoretical model.
- $\mathbf{g}_j$ is the vector of reviews for movie j.
- $\mu_j = Pr(\theta_j = 1)$ is the prior belief about the movie. It is movie specific and common to all reviewers.[3]
- $\gamma_{i,\theta} = Pr(r_{i,j} = 1|\theta_j = \theta, \mu_j)$ for $\theta = 0, 1$ are the reduced-form conditional grading probabilities.

The first step consists in maximizing the likelihood of observing the vector of reviews $\mathbf{g}_j$:

$$\max_{\{\gamma_{i,1},\gamma_{i,0}\}_{i=1}^{n},\{\mu_j\}_{j=1}^{m}} Pr(\mathbf{g}_j) = \mu_j \prod_{i=1}^{n} \gamma_{i,1}^{r_{i,j}} (1-\gamma_{i,1})^{1-r_{i,j}} + (1-\mu_j) \prod_{i=1}^{n} \gamma_{i,0}^{r_{i,j}} (1-\gamma_{i,0})^{1-r_{i,j}} \quad \text{s.t. } \gamma_{i,1} \geq \gamma_{i,0} \tag{3.1}$$

Suppose first that the movie reviewers' reputation do not affect their truthful revelation set. Then the prior on the movie would impact the grading strategies of all reviewers in the

[2] See the previous version of this paper.

[3] As in the theoretical model, we assume away heterogeneous priors. Identification does not hold in that case.

same direction, even though reviewers may be more or less biased according to their ability. Identification in this setting is proven in many papers dealing with identification of mixture models such as Allman et al. (2009). Iaryczower and Shum (2012b) use this argument to prove the identification of their structural model.

However, we have a different framework, in which the truthful revelation set of movie reviewers varies according to their initial reputation. Hence for a given prior, a highly reputed movie reviewer may be enticed to give a bad review whereas an unknown one may have incentives to give a good one[4]. We therefore need to be able to identify biases which are both prior-specific and reviewer-specific. The identification follows because we assume the variations in reputation and the variations in priors are independent.

To precise how our identification strategy works, let us detail the reduced-form grading probabilities $\gamma_{i,\theta}$ according to our structural assumptions. In the following, we denote $\underline{\mu_i}$ and $\overline{\mu_i}$ respectively the infimum and the supremum of the truthful revelation set of movie reviewer i. Although it does not appear in the notations, we allow these bounds to depend on time in the estimation as reputation may vary. We also denote $b^- = p(g = 0|s_1)$ and $b^+ = p(g = 1|s_0)$ the expert's negative and positive bias.

Table 1: Probabilities of giving a good review conditional on the true state of the world and on the prior

Prior	$\gamma_{i,1}$	$\gamma_{i,0}$
$\mu_j < \underline{\mu_i}$	$t(1 - b^-)$	$(1 - t)(1 - b^-)$
$\mu_j \in \left[\underline{\mu_i}, \overline{\mu_i}\right]$	t	$1 - t$
$\mu_j > \overline{\mu_i}$	$t + (1 - t)b^+$	$(1 - t) + tb^+$

Note that the expert always reveals her private signal if this signal confirms the prior: the bias only concerns contradictory signals. Substituting the expressions in Table 1 in the

[4]Recall that the truthful revelation set is a set in which we are sure that reviewers are not biased. We do not assume anything for what happens outside this set: reviewers may well be truthtelling so as to prove they have a high ability, or they may be biased

likelihood function yields:

$$
\begin{aligned}
\max_{\left\{t_i,b_i^-,b_i^+,\underline{\mu_i},\overline{\mu_i}\right\}_{i=1}^n,\{\mu_j\}_{j=1}^m} Pr(\mathbf{g}_j) = \\
\mu_j \prod_{i=1}^n \Bigg\{ & \mathbb{1}\left(\mu_j < \underline{\mu_i}\right) \left[(t_i(1-b_i^-))^{r_i}(1-t_i(1-b_i^-))^{1-r_i}\right] \\
&+ \mathbb{1}\left(\mu_j \in \left[\underline{\mu_i},\overline{\mu_i}\right]\right) \left[t_i^{r_i}(1-t_i)^{1-r_i}\right] \\
&+ \mathbb{1}\left(\mu_j > \overline{\mu_i}\right) \left[(t_i+(1-t_i)b_i^+)^{r_i}(1-t_i-(1-t_i)b_i^+)^{1-r_i}\right] \Bigg\} \\
+(1-\mu_j) \prod_{i=1}^n \Bigg\{ & \mathbb{1}\left(\mu_j < \underline{\mu_i}\right) \left[((1-t_i)(1-b_i^-))^{r_i}(1-(1-t_i)(1-b_i^-))^{1-r_i}\right] \\
&+ \mathbb{1}\left(\mu_j \in \left[\underline{\mu_i},\overline{\mu_i}\right]\right) \left[(1-t_i)^{r_i} t_i^{1-r_i}\right] \\
&+ \mathbb{1}\left(\mu_j > \overline{\mu_i}\right) \left[((1-t_i)+t_i b_i^+)^{r_i}(1-(1-t_i)-t_i b_i^+)^{1-r_i}\right] \Bigg\}
\end{aligned}
$$

$$
\text{s.t. } t_i \in \left[\frac{1}{2}, 1\right] \tag{3.2}
$$

The intuition for identification clearly translates into the formulation of the likelihood. Our identification stems from this partition over the set of priors: the ability t_i is identified on $\left[\underline{\mu_i}, \overline{\mu_i}\right]$, whereas the negative and positive biases, b_i^- and b_i^+, are respectively identified on $\left[0, \underline{\mu_i}\right]$ and $[\overline{\mu_i}, 1]$. Also, we can directly recover the parameters of interest of our model, t_i, b_i^-, and b_i^+, since they are related with the reduced-form conditional grading probabilities in a very simple way.

3.2 Estimation

We allow the prior μ_j to depend parametrically on movie characteristics ω_j via the following logit formulation:

$$\mu(\omega_j;\beta) = \frac{\exp(\omega_j'\beta)}{1+exp(\omega_j'\beta)} \in [0,1]$$

Elements in ω_j include information on movies available to the market prior to their release, for instance the experience of the director[5] (measured as her number of previously directed movies), the production budget, a proxy for the genre of the movie (such as MPAA ratings), whether or not the movie has been produced in the US, whether the movie is an original production, a remake, or a sequel.

Similarly, we have to estimate the initial reputation of the reviewer as it will affect her truthtelling revelation set $\left[\underline{\mu_i}, \overline{\mu_i}\right]$. More precisely, as seen in the theoretical section, $\underline{\mu_i}$ and $\overline{\mu_i}$ are both functions $\underline{\phi}$ and $\overline{\phi}$ of the prior expectation of consumers about the expert's ability, $\mathbb{E}(t_i)$. We compute these functions using the characterization of $IC_{\underline{t},\mathbb{E}(\tilde{t})}$, provided by equation 2.1. We estimate $\mathbb{E}(t_i)$ via the following logit formulation:

$$\mathbb{E}(t_i|\omega_i;\delta) = \frac{\exp(\omega_i'\delta)}{1+exp(\omega_i'\delta)} \in \left[\frac{1}{2},1\right]$$

in which ω_i is a proxy for the experts' reputation.

Since we need the reputation to be at least greater than one half[6], we impose δ to be greater than zero. Since we choose ω_i to be a proxy for the expert's reputation, we consider this assumption as innocuous. Also note that we allow the reputation as well as ω_i to vary through time.

Also, to simplify notations, we introduce the following denominations for the partition over the set of possible priors:

[5] Including the actors would be difficult as an actor's notoriety can vary greatly during her career and we would therefore need the actors' reputation for all movies in our panel.

[6] Remember that the expert's reputation is the expectation on her ability, and $\underline{t} > \frac{1}{2}$.

- $\mathbf{I_A}(\omega_i;\delta) = \left[0, \underline{\phi}(\mathbb{E}(t_i|\omega_i;\delta))\right]$: the set of priors lower than priors in the truthful revelation set. In this set, the expert can be affected by negative bias.

- $\mathbf{I_B}(\omega_i;\delta) = \left[\underline{\phi}(\mathbb{E}(t_i|\omega_i;\delta)), \overline{\phi}(\mathbb{E}(t_i|\omega_i;\delta))\right]$: the truthful revelation set.

- $\mathbf{I_C}(\omega_i;\delta) = \left[0, \overline{\phi}(\mathbb{E}(t_i|\omega_i;\delta))\right]$: the set of priors greater than priors in the truthful revelation set. In this set, the expert can be affected by positive bias.

All these sets depend on the reputation of each expert via δ and ω_i.

$\left\{t_i, b_i^-, b_i^+\right\}_{i=1}^n$ are reviewer-specific and estimated using only the distribution of reviews.

All the parameters are estimated by maximizing the following likelihood function:

$$
\begin{aligned}
&\max_{\left\{t_i, b_i^-, b_i^+\right\}_{i=1}^n, \beta, \delta} \\
&\sum_j \log\Bigg[\mu(\omega_j;\beta)\prod_{i=1}^n\Bigg\{\mathbb{1}\Big(\mu(\omega_j;\beta)\in\mathbf{I_A}(\omega_i;\delta)\Big)\left[(t_i(1-b_i^-))^{r_i}(1-t_i+t_ib_i^-)^{1-r_i}\right] \\
&\qquad +\mathbb{1}\Big(\mu(\omega_j;\beta)\in\mathbf{I_B}(\omega_i;\delta)\Big)\left[t_i^{r_i}(1-t_i)^{1-r_i}\right] \\
&\qquad +\mathbb{1}\Big(\mu(\omega_j;\beta)\in\mathbf{I_C}(\omega_i;\delta)\Big)\left[(t_i+(1-t_i)b_i^+)^{r_i}((1-t_i)(1-b_i^+))^{1-r_i}\right]\Bigg\} \\
&+(1-\mu(\omega_j;\beta))\prod_{i=1}^n\Bigg\{\mathbb{1}\Big(\mu(\omega_j;\beta)\in\mathbf{I_A}(\omega_i;\delta)\Big)\left[((1-t_i)(1-b_i^-))^{r_i}(t_i+(1-t_i)b_i^-)^{1-r_i}\right] \\
&\qquad +\mathbb{1}\Big(\mu(\omega_j;\beta)\in\mathbf{I_B}(\omega_i;\delta)\Big)\left[(1-t_i)^{r_i}t_i^{1-r_i}\right] \\
&\qquad +\mathbb{1}\Big(\mu(\omega_j;\beta)\in\mathbf{I_C}(\omega_i;\delta)\Big)\left[((1-t_i)+t_ib_i^+)^{r_i}(t_i(1-b_i^+))^{1-r_i}\right]\Bigg\}\Bigg]
\end{aligned}
$$

$$
\text{s.t. } t_i \in \left[\frac{1}{2}, 1\right], \quad b_i^+, b_i^- \in [0,1], \delta \geq 0 \tag{3.3}
$$

In appendix B, we provide a more robust estimation strategy taking into account potential conflicts of interest between the reviewer and the production company of the movie.

We conclude this section by stating two assumptions we used in our estimation:

Assumption 2. $\underline{t} = 0.55$

Assumption 3. *The precision of the aggregate signal X is linear in the intermediate posterior of the consumers, more precisely:* $\tau(\nu) = \frac{1+\nu}{2}$.

By definition, the truthful revelation set is the set of incentive compatible priors corresponding to the lowest ability $\underline{t} \geq \frac{1}{2}$. Since we do not know the value of the lowest ability in reality, a robust approach would be to choose $\underline{t}$ as close to one half as possible. Indeed, by proposition 2, the chosen truthful revelation set would lie within the actual one. However, the truthful revelation set converges to a line when $\underline{t}$ tends toward one half, making it impossible to identify the ability[7]. We therefore choose $\underline{t} = 0.55$ in our estimations. We feel this value is close enough to one half to be robust, a feeling confirmed by our results: all our estimates of the reviewers' ability are well above this threshold.

Assumption 3 is a simplifying assumption restricting the way consumers react to the reviews. In the future, we plan to estimate a parametric formulation of $\tau(\nu) = \frac{1+\nu^{\alpha}}{2}$ with $\alpha \in \mathbb{R}^{+}$.

A graph depicting the truthful revelation set used for our estimations can be found in appendix D. The shape of this set depends on these two previous assumptions.

4 Presentation of the Data

To conduct our empirical analysis, we use movie ratings by professional movie reviewers published in the rottentomatoes.com website. This website gathers reviews from the most prominent movie reviewers in North America. The dataset forms a panel whose two dimensions are the movies and the reviewers. Movies are either judged *fresh* or *rotten*, a binary

[7]When the truthful revelation set is a line, the probability that an observation falls within is zero.

structure completely in line with our theoretical model and structural technique of estimation. The reviews provided by rottentomatoes correspond to rescaled grades of the original reviews' numerical or alphabetical scores. When the original review is associated with no grade, or the grade is average (e.g. 3 stars out of 5), rottentomatoes assigns a rating based on the tone of the review. We restrict our dataset to ratings by reviewers qualified as *top critics* by rottentomatoes on movies released in the US between 1990 and 2013. The reviewers in our sample are therefore professional reviewers driven by career concerns.

Each review contains the newspaper or TV show in which it was given, the name of the reviewer, and the date of the publication. We only keep reviews which are close in time so as to ensure that competition is indeed simultaneous and no herding behavior occurs.

We combine these reviews with data on movies' characteristics, collected on the imdb.com website (the Internet Movie DataBase). For each movie, the dataset includes the official release date, the total box-office revenues, the production budget, the number of screens on which the movie was run during the opening weekend, the MPAA rating, the genre and whether or not the movie was produced in the US. In addition, for each movie, the dataset contains the name of the director(s) and the number of films they directed in the past.

Finally, as we need to control for reputation, we include some reviewers' characteristics, namely an index summarizing the number of monthly searches on Google provided by the website Google Trends, and some charcteristics of the newspapers or TV shows for which they gave their reviews. We also include a dummy variable taking value 1 if a potential conflict of interest exists between the newspaper and the production company of the movie.

Our complete sample contains a total of 118,208 reviews over 5,578 movies by 1,242 reviewers. The variables in our dataset are extensively detailed in appendix C.

In the following, we describe more precisely how we build our dataset.

4.1 Movie reviewers

Rottentomatoes defines its *top critic* category on the basis of the size of their audience[8].

785 out of the 1,242 reviewers in our sample published less than 10 movie reviews, which represents about 63% of our population of reviewers. 44 of them individually published more than 850 reviews, our most prolific reviewer being the famous Roger Ebert with 3,277 reviews. Since we estimate experts' individual abilities and, to do so, need a consequent number of reviews, we focus on these reviewers[9].

To approximate the reputation of these reviewers, we choose the reputation proxy ω_i to be an index representing the number of monthly Google searches for each reviewer. This index is provided by the website Google Trends and takes values between 0 and 100, 100 being attributed to the point in time when the popularity of the reviewer is maximal. When comparing the popularity of two reviewers, the returned values are computed relative to the maximum number of searches for either of them over the period. Since the most popular reviewer in our sample, Roger Ebert, is clearly an outlier in terms of reputation and we do want to observe some variations between reviewers' reputation, we choose a moderately popular reviewer, Peter Travers, as our benchmark and compare all the others to him.

To take into account the experts' reputation, we have to disregard nine reviewers as they have highly popular homonyms and we cannot disentangle their reputation from the reputation of the celebrity they share their name with. For instance, we have to exclude a Roger Moore, who clearly never interpreted a famous British secret service agent[10]. We end up with 35 reviewers, whose list is found in appendix C. Overall these reviewers produced

[8] According to Rottentomatoe, "To be considered for Top Critics designation, a critic must be published at a print publication in the top 10% of circulation, employed as a film critic at a national broadcast outlet for no less than five years, or employed as a film critic for an editorial-based website with over 1.5 million monthly unique visitors for a minimum of three years. A Top Critic may also be recognized as such based on their influence, reach, reputation, and/or quality of writing, as determined by Rotten Tomatoes staff."

[9] The threshold of 850 reviews is arbitrary: it gives the maximum number of movie reviewers for an acceptable amount of reviews.

[10] Our systematic rule to determine whether or not a reviewer has a famous homonym is typing their name in a web search engine and checking that all first results concern a movie reviewer.

48,278 reviews, which represents about 41% of our observations.

We also observe the medium through which the reviewer gave her review and the precise date of the review. Reviewers in our sample publish in various kinds of media, from national TV and radio shows to local newspapers. We are able to identify whether or not the medium is addressed to a general or a specialized audience, whether or not its coverage is nationwide or statewide. The date of the review allows us to ensure that we analyze a simultaneous competition game. Since Google Trends data only start in 2004, we only consider reviews given from 2004 to 2013. We also exclude movies released prior to 2003 as some movies of the 1990s are given late reviews for their DVD release ten years later[11]. We consider that these reviews could be biased by some herding behaviors which do not enter into the scope of our analysis. Our final dataset shows an average standard deviation of 39 days for the dates of the reviews. We aim to reduce this standard deviation by excluding late reviews in future estimations.

We finally exclude all movies with less than four reviews because: (a) we need variations in the reviews to identify the prior of the movie, (b) we want movies as homogeneous as possible in terms of number of reviews to lift the issue of a potential selection bias from reviewers. Our final sample contains 30,531 reviews over 2,413 movies.

4.2 Description of the Movies

Our dataset consists of movies released in the US between 1990 and 2013 which have been granted a rating by the Motion Picture Association of America. The MPAA ratings are guidelines given to parents on the contents of movies: they consist in the five following different grades G, PG, PG-13, R, and NC-17. Although these ratings are not central to our analysis, we choose to focus on MPAA rated movies because they represent a large part of all movies released in the US. Indeed, the six main production companies are part of the

[11]We choose to include 2003 to keep the movies released in late 2003 and reviewed in early 2004.

MPAA and must therefore submit their products to the rating system.

Table 2 presents some descriptive statistics on the movies in our sample. As shown by the converted budget in today's US$ and the width of the release as the number of screens in the opening weekend, our sample does not only features blockbusters but also more independent movies. Indeed, the average production budget for a wide release in the USA is $66 million[12] which is twice our sample average. Also, our median movie is released on only 179 screens in the opening week, which is below the cut-off value above which you can consider a movie as a wide release: 600 screens according to Einav (2007). That does not necessarily mean that lots of movies in our sample are small confidential productions, also known as movies in limited release, but more probably some of our movies have been first released in a few cities to build a reputation and then have been displayed on more screens. These movies are called *platform release*. The US gross profit expressed in today's US$ also shows a diversity in the way movies have succeeded, from big hits to big failures, with a relatively high standard deviation. # weeks stands for the total number of weeks during which the movie has been screened on some theater in the US. On average a theater runs a movie from 6 to 8 weeks (see Einav (2007)), but some might run it longer, so our median total number of weeks of 11 seems consistent with this previous observation. We also include the number of previous movies by the same director, which should influence the prior on the movie, and is concentrated around low values in our dataset, with a long thin tail.

4.3 Media Outlets

We conclude the description of our data by detailing the potential conflicts of interest we find in our sample. We follow DellaVigna and Kennedy (2011) and create a dummy variable equal to 1 if the production company of the movie and the medium of the review belong to the same media outlet at the time of the review. Since this question is not central to our

[12] source: www.the-numbers.com

Table 2: Some descriptive statistics on the movies

	B.O. US (M$)	# weeks	# screens 1st week	Budget (M$)	#Previous Movies
# obs.	4821	4563	4312	3556	5578
Median	9.7	11	179	23	3
Mean	38.9	12.15	1195	39	5.33
St. Dev.	68.7	9.32	1342	45	7.18
Highest	Titanic: 951	Roving Mars: 167	The Dark Knight Rises: 4404	Pirates of the Caribbean 3: 335	Chunhyangdyun (Im, Kwon-taek): 96
2nd	Avatar: 821	Deep Sea: 165	Iron Man 2: 4380	Titanic: 289	Éloge de l'amour (Godard, Jean-Luc): 74
3rd	The Avengers: 629	Aliens of the Deep: 142	Harry Potter and the Deathly Hallows - Part 2: 4375	Spider-Man 3: 288	Kakushi ken oni no tsume (Yamada, Yôji): 72

paper but included to show that our results are not driven by the wrong kind of bias, we do not provide a description as detailed as in their paper.

We identify a new potential source of conflicts of interest: the Disney Media Group which controls the Ebert & Roeper TV show and studios such as Miramax and Walt Disney Pictures.

Overall, only 1.4% of our observations are concerned with these conflicts of interest. All possible conflicts of interest that we are aware of in our sample are detailed in appendix C. Note that when reviewers express themselves on several media, we only consider the reviews in media in the same group as the production companies to be problematic.

5 Results

We now turn to the results of our empirical analysis. Table 3 and Table 4 respectively present the movie-specific estimates and a summary of the reviewer-specific estimates under two alternative specifications. Compared to specification (I), specification (II) includes the budget in the determination of the prior and takes into account potential conflicts of interest. Table 12 in appendix D gives the exhaustive list of the reviewer-specific parameters' estimates.

The prior depends positively and significantly on a movie being a foreign production. Our interpretation is that low quality foreign films are in general not successful enough to be released in the US. When facing a foreign movie, consumers rationally expect it to be of high quality. Remakes and sequels also impact negatively the prior. The result on sequels is in line with an empirical observation that consumers give significantly lower grades to sequels: in general the sequel must be of lower quality than the original work, hence the lower prior. The same story can apply to remakes, although the coefficient in this case is less significant. As expected, the director's experience impacts positively and significantly the prior on the movie. Compared to a PG-13 MPAA rating, G and R ratings impact positively the prior whereas PG and NC-17 are not significant. We use these ratings as controls for the content of the movies. Surprisingly, the budget has no strong impact on the prior. Our interpretation is that the gains from a larger budget, i.e. better actors, special effects, etc., might be counterbalanced by a loss in originality and quality from big productions.

The coefficient on the Google search index is quite small and suggests that consumers' manipulation by movie reviewers is limited. With such a coefficient, only Roger Ebert would have had a real shot at decreasing demand to maintain his reputation, but as the estimates of his negative bias suggests, he did not use this mechanism. However, this result means the model with endogenous realization of the state of the world is a better fit to the data

Table 3: ML estimates for the prior, reputation, and conflicts of interest — Specification (II) includes the budget and conflicts of interest

	(I)		(II)	
	Coeff.	Bootstrap SE	Coeff	Bootstrap SE
Movie Specific, β:				
Constant	1.249	(0.040)	1.124	(0.004)
Origin: USA	-1.494	(0.038)	-1.440	(0.016)
Origin: co-production USA	-1.539	(0.074)	-1.554	(0.083)
Remake	-0.501	(0.247)	-0.439	(0.228)
Sequel	-0.496	(0.053)	-0.432	(0.115)
Number of director's previous films	0.016	(0.000)	0.016	(0.000)
G rating	0.949	(0.173)	0.988	(0.332)
PG rating	0.087	(0.076)	0.470	(0.191)
R rating	0.484	(0.043)	0.530	(0.040)
NC-17 rating	0.690	(9.160)	0.650	(4.117)
log budget			0.000	(0.000)
Reputation, δ:				
Google Search Index	5.9×10^{-4}	(0.000)	5.9×10^{-4}	(0.000)
Conflict of interest, b^c:				
Average Bias			0.055	(0.034)
# Observations:	30440		22674	

Notes: Bootstrap Standard Errors are computed on 100 iterations for (I) and 500 for (II)

than the model in which we learn the state of the world at the end of the game. Indeed, the estimates do suggest that the truthful revelation set is shifted towards high prior movies if the reputation is high enough. If experts were not able to manipulate demand, $\hat{\hat{\delta}}$ would have been equal to zero.

The small amplitude of the effect might be an artifact of two assumptions: (a) the linearity of the precision of the aggregate signal, (b) the implicit assumption that movie reviewers operate in autarkic markets. To check the robustness of our findings, we can (a) estimate the precision of the aggregate signal as a function of the prior and the reviews, (b) estimate a model in which movie reviewing is part of a competitive market[13].

Finally, we find a positive bias due to conflict of interest. This bias is quite large since reviewers turn a bad review into a good one 5% of the time when reviewing a movie from the same media outlet. It is significant at an 11% level.

All our estimates are robust to the change in specification from (I) to (II). This is in line with the fact that the budget has no effect on the prior and that conflicts of interest should also be independent from the movie prior.

Table 4 summarizes our findings on the reviewer-specific parameters, i.e. their ability, negative bias, and positive bias. First, all the estimates are robust to the change in specifications: the number of instances in which we have a conflict of interest is probably too small to bias the estimates of b^- and b^+ in (I). The average ability of experts in our sample is quite high, with movie reviewers being able to correctly observe the quality of a movie 78% of the time. The average positive and negative biases are significant with reviewers reporting a negative opinion after receiving a positive signal and vice versa respectively 8% and 15% of the time.

The highest ability reviewer, Robert Denerstein, is very efficient, with a precision of 90%

[13] In that case, the fact that we exclude movies with less than four reviews necessarily leads to an underestimation of the feasibility of manipulation.

Table 4: Summary of ML estimates for reviewer-specific parameters — Specification (II) includes the budget and conflicts of interest

		(I)			(II)		
		t	b^-	b^+	t	b^-	b^+
Median		0.78	0.04	0.11	0.77	0.08	0.11
Mean		0.77	0.08	0.15	0.77	0.011	0.13
St. Dev.		0.06	0.11	0.13	0.06	0.011	0.13
Highest ability	Robert Denerstein:	0.898	0.107	0.001	0.898	0.146	0.013
		(0.009)	(0.035)	(0.034)	(0.021)	(0.062)	(0.042)
Lowest ability	Kyle Smith:	0.619	0.345	0.000	0.654	0.363	0.000
		(0.015)	(0.047)	(0.000)	(0.026)	(0.065)	(0.000)
# Observations:			30440			22674	

Notes: Bootstrap Standard Errors are computed on 100 iterations for (I) and 500 for (II)

and no bias on high prior movies. However, for low prior movies, he will send a bad review after observing a positive signal 11% of the time. Kyle Smith, the lowest ability reviewer, only observe the accurate signal 62% of the time and has a strong negative bias of 34%.

A surprising feature of our results is that some reviewers are apparently strongly polarized, i.e. they are either prone to negative criticism or to positive criticism. For instance, Liam Lacey has a strong negative bias of 41%. But he is truthtelling for movies with positive priors. On the opposite, Joe Baltake is completely truthtelling for low prior movies, but has a strong positive bias of 43% for high prior movies. Whether or not these differences are due to different personal tastes towards negative or positive criticism, indications that reviewers are heterogeneous in their abilities to recognize high and low qualities, or mere features of the strategies played at equilibrium, is to this point unknown to us.

Finally, it does seem that more able experts are less prone to bias: we find a slightly negative relationship between bias and ability. However, having observations on only 35 reviewers, this relationship is not significant.

6 Out-of-sample Predictions

In this section, we test the predictive power of our model. We find that the predicted distribution of reviews is very close to the distribution of reviews observed in the data. Our model also better fits out-of-sample data than two alternative models: one ruling out strategic biases and one assuming that all experts are equally well informed and that their ability is common knowledge.

6.1 Testing Predictive Power

Applying our model's estimates to new data, i.e. new movies, we can recover the priors on their quality but also the ex-ante probability, i.e. prior to observing the private signal, that a particular reviewer gives a good or bad review to a movie. However, we only observe the realization of this ex-ante distribution, which is the final review. Testing the predictive power of our model is therefore difficult since we cannot directly compare the estimated and realized outcomes.

To overcome this difficulty, we group together observations, i.e. couples of movie and movie reviewer, for which our model predicts similar probabilities of good reviews. Within each group, we then compute the actual proportion of good reviews which corresponds to the observed probability of a good review conditional on belonging to the group. With this method, we cannot generally conclude that a model with similar predicted and observed probabilities is a good predictor[14], however we can rule out models whose predicted probabilities differ widely from the actual ones.

We provide an additional proof of the relevance of our model by comparing its predictive power to the one of alternative models. To achieve this goal we compute the confusion

[14] Especially if the rule allocating an observation to a group is too coarse. For instance, if we take all observations in one group, the average review in the main data is a pretty good estimator of the average review in the out-of-sample data.

matrix of our model: for each movie-movie reviewer couple in our out-of-sample data, we draw a realization of the review from our estimated ex-ante distribution of giving a good or a bad review. We then compare this realization to the actual review given by the reviewer. If they are similar, we say our model classified the observation correctly. Repeating that several times, we can obtain an average classification ratio. Doing that same operation for alternative models, the best one has the highest average classification ratio.

6.1.1 Out-of-sample Data

We randomly split our main data set used in section 5 and estimate our model on 80% of the movies in our sample. The 20% remaining, which form our out-of-sample data set, are used to test the out-of-sample predictive power of our model. Out of the 1661 movies and 22674 reviews in the main data set, the out-of-sample data set contains 351 movies for a total of 4792 reviews. This out-of-sample dataset needs to be representative. This is in theory ensured by the fact that it is contructed randomly. As additional evidence, table 5 shows that moments of several variables are similar between our main data set and the out-of-sample data set.

6.1.2 Model Specifications

Our main econometric model is model (II) from section 5. It includes the budget in the prior of the movies and potential conflicts of interest which can arise when the movie's production company and the reviewer's newspaper or show belong to the same production company. To carry out out-of-sample predictions, we need to estimate this model again using our restricted data set. Estimates are reported in tables 13 and 14 of appendix D and are remarkably close to the ones using the full data set.

We compare this model to alternative theories. The first one, called model (III), assumes that reviewers have a private ability and observe a private signal but are always truthful.

Table 5: Comparison of Moments between the Main Data Set and the Data Set Used for Out-of-sample Predictions

	Main	Out-of-sample
Avg. Budget (M$)	43.7	41.9
Director's Avg. Number of Previous Movies	6.50	6.95
Freq. of US productions	57.3%	57.0 %
Avg. Number of Reviews	13.65	13.65
Freq. of *Fresh* Reviews	49%	48%
Proportion of movies released in:		
2003	2.1%	4.3%
2004	11.3%	12.0%
2005	12.5%	10.3%
2006	12.3%	12.3%
2007	11.6%	10.5%
2008	11.4%	11.7%
2009	10.7%	10.3%
2010	9.3%	10.8%
2011	10.3%	8.8%
2012	7.2%	7.4%
2013	1.4%	1.7%

Model (III) is therefore a model excluding strategic biases. We estimate the parameters of this model by maximizing the following likelihood:

$$\max_{\{t_i\}_{i=1}^n,\beta} \quad \sum_j \log\left[\mu(\omega_j;\beta)\prod_{i=1}^n\left[t_i^{r_i}(1-t_i)^{1-r_i}\right]+(1-\mu(\omega_j;\beta))\prod_{i=1}^n\left[(1-t_i)^{r_i}t_i^{1-r_i}\right]\right] \text{ s.t. } t_i\in\left[\frac{1}{2},1\right]$$

The second one, called model (IV), assumes that reviewers observe a private signal which depends solely on the publicly observable prior plus an i.i.d. noise, which is independent from their ability. In this model, reviewers also reveal truthfully their signal since their payoff does not depend on their reputation for being well informed. Consequently, reviews follow the prior distribution, which we can estimate using a simple logit approach.

Estimates for the two alternative models are reported in tables 13 and 14 of appendix D.

6.2 Out-of-sample Fit and Classification Power

Our first approach consists in computing the ex-ante distribution of good and bad reviews as predicted by our model (II) for each observation in our out-of-sample data and grouping together the observations for which our model predicts a similar distribution of reviews. We then compare the predicted probabilities of good and bad reviews with the frequencies observed in each groups. Table 6 and figure 3 show that our model performs quite well according to this criterion. In table 6, we group predicted probabilities of good reviews by increments of 10%. For each group with a non-negligible amount of observations except one, the differences between the predicted and observed probabilities of a good review are less than 5%. The only exception arises in $[0.5, 0.6]$, in which the difference is around 8%.

In figure 3, out-of-sample observations are sorted horizontally according to their predicted probabilities of good reviews. The red curve represents a non-parametric fit of the actual proportion of good reviews. Groups here are defined by the kernel of the fit and the closer the red curve is to the 45-degree line, the better our model is at predicting reviews.

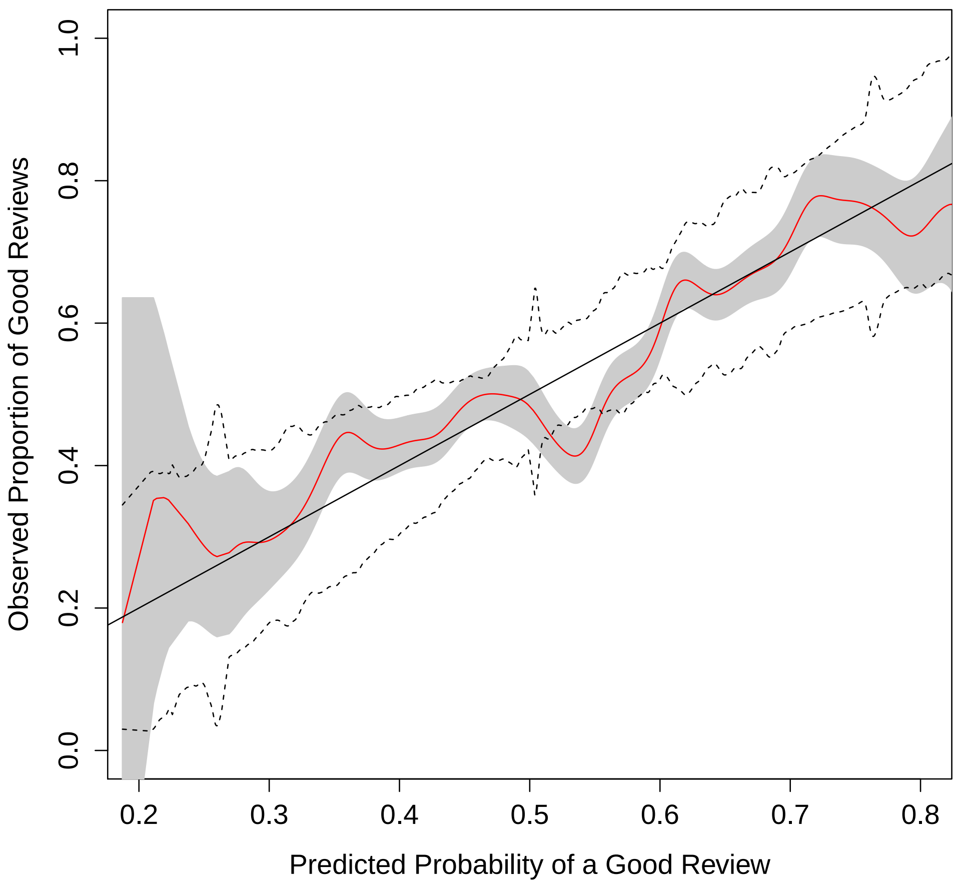

Figure 3: Observed Proportion of Good Reviews given their Predicted Probability — The red line displays the non-parametric fit of the observed proportion of good reviews. The shaded area represents the 95% confidence interval of this fit. The black plain line is the 45 degree line. The black dotted line represents the 95% confidence interval for the predicted distribution.

Table 6: Predicted Probabilities of Good Reviews and their Observed Frequency in the Data

$\widehat{P(r_{i,j} = 1)}$ $\in$	# Obser.	Avg. Predicted Probability of a Good Review	Observed Freq. of Good Reviews
$[0, 0.1]$	0	—	—
$[0.1, 0.2]$	1	18.73%	0
$[0.2, 0.3]$	88	26.37%	29.55%
$[0.3, 0.4]$	678	35.80%	39.38%
$[0.4, 0.5]$	1319	44.70%	47.38%
$[0.5, 0.6]$	1194	55.27%	47.32%
$[0.6, 0.7]$	1112	64.91%	66.19%
$[0.7, 0.8]$	356	74.70%	76.97%
$[0.8, 0.9]$	44	81.40%	77.27%
$[0.9, 1]$	0	—	—

Predicted

$$\text{Observed} \quad \begin{array}{c} \\ 0 \\ 1 \end{array}\!\!\begin{array}{c} \begin{array}{cc} 0 & \quad 1 \end{array} \\ \begin{bmatrix} 1134.5 & 1130.5 \\ 1117 & 1410 \end{bmatrix} \end{array} \qquad \begin{array}{c} \\ 0 \\ 1 \end{array}\!\!\begin{array}{c} \begin{array}{cc} 0 & \quad 1 \end{array} \\ \begin{bmatrix} 1174.2 & 1090.8 \\ 1231.8 & 1295.2 \end{bmatrix} \end{array} \qquad \begin{array}{c} \\ 0 \\ 1 \end{array}\!\!\begin{array}{c} \begin{array}{cc} 0 & \quad 1 \end{array} \\ \begin{bmatrix} 1085 & 1180 \\ 1124.1 & 1402.9 \end{bmatrix} \end{array}$$

Model (II) Model (III) Model (IV)

Figure 4: Confusion Matrices Averaged over 100 Repetitions for Models (II), (III), and (IV) — The first row and column of each matrix respectively represent observed and predicted bad reviews. The second row and column are similarly defined for good reviews.

We now compare confusion matrices for our model and for alternative models, namely (III) and (IV). Elements of the matrix are the number of observations corresponding to the predicted and observed outcomes. The overall classification ratio is similarly defined as the number of observations properly predicted over the total number of observations. A high classification ratio is indicative of a high predictive power. Figure 4 and table 7 respectively presents the confusion matrices and classification ratios averaged over 100 repetitions for the three models under study.

Table 7: Classification Ratios Averaged over 100 Repetitions for Models (II), (III), and (IV)

	(II)	(III)	(IV)
Overall Classification Ratio	0.531	0.515	0.519
Classification Ratio on Good Reviews	0.558	0.513	0.555
Classification Ratio on Bad Reviews	0.501	0.518	0.479

Classification ratios are generally quite low. This is to be expected given that for a large part of our observations, our models predict probabilities of good reviews around one half. In this situation, the model misclassifies an observation one half of the time. This tool is therefore relevant only insofar as it allows us to compare several models. Despite some issues with the classification of bad reviews, with an overall classification ratio of 0.53, our model performs better than the alternative specifications. In conjonction with the good fit of predicted probabilities of good reviews to observed ones, this is a good indicator that models of expertise with reputational concerns and strategic biases best describe the behavior of movie reviewers.

7 Conclusion

We estimate the strategic incentives of experts to send biased recommendations using movie reviews. In our model, experts want to maximize their reputation as good predictors of the state of the world. Bias in this case can take two forms: experts can disregard contradictory signals because they are noisy, and they can send negative recommendations to discourage demand and hinder the update on their reputation. In this situation, expert bias depends both on the prior on the state of the world and on the expert through her ability and reputation.

To tackle these issues, and the fact that the state of the world, abilities, and private signals are not observable, we introduce new identification and estimation strategies. The

prior is identified by the fact that signals, and therefore reviews, are correlated through the unobservable true quality. The experts' abilities are identified by the fact that experts are unbiased on a subset of the priors. Outside this subset, the difference between the actual distribution of the expert's reviews and the one generated by a truthtelling expert with the same ability gives us the bias. The estimation is a straightforward one-step process, which allows us to recover the determinants of the prior, the relative impact of the reputation, any bias due to a conflict of interest, and the ability and biases of each expert in our sample.

We find strong variations in movie reviewers' abilities and biases. The average negative and positive biases in our sample are strong and significant, with information manipulation occuring on average 10% of the time. Also, we find a small but significant impact of reputation validating models with endogenous realization of the state of the world at the expense of models in which it is revealed.

We devote to future research the introduction of heterogeneous abilities to recognize high quality and low quality movies, the case of competitive review markets, the estimation of the precision of the aggregate signal, and a fully structural estimation of the reputation.

References

Elizabeth S Allman, Catherine Matias, and John A Rhodes. Identifiability of parameters in latent structure models with many observed variables. *The Annals of Statistics*, 37(6A): 3099–3132, 2009.

Suman Basuroy and S Abraham Avri Ravid. How relevant are experts in the internet age? evidence from the motion pictures industry? 2013.

Suman Basuroy, Subimal Chatterjee, and S Abraham Ravid. How critical are critical reviews?

the box office effects of film critics, star power, and budgets. *Journal of Marketing*, pages 103–117, 2003.

Peter Boatwright, Suman Basuroy, and Wagner Kamakura. Reviewing the reviewers: The impact of individual film critics on box office performance. *Quantitative Marketing and Economics*, 5(4):401–425, 2007.

F. Camara and N. Dupuis. Avoiding judgement by recommending inaction: Beliefs manipulation and reputational concerns. 2013.

S. DellaVigna and A. Kennedy. Does media concentration lead to biased coverage? evidence from movie reviews. 2011.

Loretti Dobrescu, Michael Luca, and Alberto Motta. What makes a critic tick? connected authors and the determinants of book reviews. 2012.

L. Einav. Seasonality in the us motion picture industry. *The Rand journal of economics*, 38 (1):127–145, 2007.

J. Eliashberg and S.M. Shugan. Film critics: Influencers or predictors? *The Journal of Marketing*, pages 68–78, 1997. ISSN 0022-2429.

M. Gentzkow and J.M. Shapiro. Media bias and reputation. *Journal of Political Economy*, 114(2):280–316, April 2006.

Matias Iaryczower and Matthew Shum. Money in judicial politics: Individual contributions and collective decisions. *Available at SSRN 1998895*, 2012a.

Matias Iaryczower and Matthew Shum. The value of information in the court: Get it right, keep it tight. *The American Economic Review*, 102(1):202–237, 2012b.

Matias Iaryczower, Garrett Lewis, and Matthew Shum. To elect or to appoint? bias, information, and responsiveness of bureaucrats and politicians. *Journal of Public Economics*, 2012a.

Matias Iaryczower, Xiaoxia Shi, and Matthew Shum. Words get in the way: The effect of deliberation in collective decision-making. 2012b.

Beatriz Mariano. Market power and reputational concerns in the ratings industry. *Journal of Banking & Finance*, 2012.

M. Ottaviani and P.N. Sørensen. Reputational cheap talk. *The Rand journal of economics*, 37(1):155–175, 2006.

D.A. Reinstein and C.M. Snyder. The influence of expert reviews on consumer demand for experience goods: A case study of movie critics. *The journal of industrial economics*, 53 (1):27–51, 2005. ISSN 1467-6451.

Appendix

Appendix A provides the proofs of the propositions in section 2. Appendix B precises the estimation strategy in the presence of conflicts of interest. Appendix C gives more details about our dataset. Finally, appendix D completes the exposition of our results with the exhaustive list of estimates for the reviewer-specific parameters.

A Proofs of the Propositions

A.1 Proof of Proposition 1

We need to characterize the experts' incentive to be truthtelling when the market thinks she is. Therefore, we assume in this proof that the belief of the market on the strategy of the expert is $\tilde{\sigma} = \sigma^T$. Given this, an expert sending a recommendation r_i is perceived by the consumers as having private infomation $\tilde{s}_i$. The truthtelling incentive is satisfied as long as revealing s_i yields a higher expected utility:

$$\begin{aligned} & \mathbb{E}_f(u(\tilde{t})|s_i, \tilde{s}_i, t) \geq \mathbb{E}_f(u(\tilde{t})|s_i, \tilde{s}_{-i}, t) \\ \Leftrightarrow & \int_{\underline{t}}^{\bar{t}} u(\tilde{t}) dF(\tilde{t}|s_i, \tilde{s}_i, t) \geq \int_{\underline{t}}^{\bar{t}} u(\tilde{t}) dF(\tilde{t}|s_i, \tilde{s}_{-i}, t) \end{aligned} \tag{A.1}$$

in which $F(\tilde{t}|s_i, \tilde{s}, t)$ is the expert's expectation of her reputation update by consumers after sending a recommendation $r = \tilde{s}$.

Denoting f the density associated to F, we get:

$$f(\tilde{t}|s_i, \tilde{s}, t) = f(\tilde{t}|\tilde{s}, \theta_i) p^e(\tilde{\theta}_i|s_i, \tilde{s}, t, \mathbb{E}(\tilde{t})) + f(\tilde{t}|\tilde{s}, \theta_{-i}) p^e(\tilde{\theta}_{-i}|s_i, \tilde{s}, t, \mathbb{E}(\tilde{t}))$$

In this equation, $p^e(\tilde{\theta}|s_i, \tilde{s}, t, \mathbb{E}(\tilde{t}))$ is the expert's expectation on the consumers' belief on the state of the world at the end of the game. Notice that this expectation does not

depend on the whole prior belief on the expert's ability $f(\tilde{t})$ but only on its first moment $\mathbb{E}(\tilde{t})$. Indeed, $f(\tilde{t})$ only affect $p^e(\tilde{\theta}|s_i, \tilde{s}, t)$ through the consumers' intermediate posterior after observing the recommendation of the expert: $p(\theta_i|\tilde{s}_i) = \frac{\mathbb{E}(\tilde{t})p(\theta_i)}{\mathbb{E}(\tilde{t})p(\theta_i)+(1-\mathbb{E}(\tilde{t}))p(\theta_{-i})}$.

The reputation conditional on $\tilde{s}$ and θ_i is updated as follows:

$$\begin{cases} f(\tilde{t}|\tilde{s}_i, \theta_i) = \frac{p(\tilde{s}_i|\theta_i,\tilde{t})f(\tilde{t}|\theta_i)}{\int_{\underline{t}}^{\bar{t}} p(\tilde{s}_i|\theta_i,\tilde{t})f(\tilde{t}|\theta_i)d\tilde{t}} = \frac{\tilde{t}f(\tilde{t})}{\mathbb{E}(\tilde{t})} \\ f(\tilde{t}|\tilde{s}_{-i}, \theta_i) = \frac{(1-\tilde{t})f(\tilde{t})}{1-\mathbb{E}(\tilde{t})} \end{cases}$$

Substituting these expressions in inequality A.1 yields:

$$\begin{aligned} \int_{\underline{t}}^{\bar{t}} u(\tilde{t}) \left(\frac{\tilde{t}f(\tilde{t})}{\mathbb{E}(\tilde{t})} - \frac{(1-\tilde{t})f(\tilde{t})}{1-\mathbb{E}(\tilde{t})} \right) \left[p^e(\tilde{\theta}_i|s_i, \tilde{s}_i, t, \mathbb{E}(\tilde{t})) - p^e(\tilde{\theta}_{-i}|s_i, \tilde{s}_{-i}, t, \mathbb{E}(\tilde{t})) \right] d\tilde{t} \geq 0 \\ \Leftrightarrow \frac{p^e(\tilde{\theta}_i|s_i, \tilde{s}_i, t, \mathbb{E}(\tilde{t})) - p^e(\tilde{\theta}_{-i}|s_i, \tilde{s}_{-i}, t, \mathbb{E}(\tilde{t}))}{\mathbb{E}(\tilde{t})(1-\mathbb{E}(\tilde{t}))} \int_{\underline{t}}^{\bar{t}} \tilde{t} \left(\tilde{t} - \mathbb{E}(\tilde{t}) \right) f(\tilde{t}) d\tilde{t} \geq 0 \\ \Leftrightarrow p^e(\tilde{\theta}_i|s_i, \tilde{s}_i, t, \mathbb{E}(\tilde{t})) - p^e(\tilde{\theta}_{-i}|s_i, \tilde{s}_{-i}, t, \mathbb{E}(\tilde{t})) \geq 0 \end{aligned}$$

in which the second inequality follows from $u(\tilde{t}) = \tilde{t}$ and the last one is given by:

$$\int_{\underline{t}}^{\bar{t}} \tilde{t} \left(\tilde{t} - \mathbb{E}(\tilde{t}) \right) f(\tilde{t}) d\tilde{t} = \int_{\underline{t}}^{\bar{t}} \tilde{t}^2 f(\tilde{t}) d\tilde{t} - \mathbb{E}(\tilde{t}) \int_{\underline{t}}^{\bar{t}} \tilde{t} f(\tilde{t}) d\tilde{t} = \mathbb{E}(\tilde{t}^2) - \mathbb{E}(\tilde{t})^2 \geq 0$$

A.2 Proof of Corollary 1

When $\tau(\nu) = 1 \, \forall \nu$, $p^e(\tilde{\theta}|s, \tilde{s}, t, \mathbb{E}(\tilde{t})) = p(\theta|s, t)$, therefore the characterization is equivalent to $p(\theta_i|s_i, t) \geq p(\theta_{-i}|s_i, t) \Leftrightarrow \mu \in [1-t, t]$.

A.3 Proof of Proposition 2

Let us first rewrite the characterization of the incentive compatibility condition. To simplify notations, we omit t and $\mathbb{E}(\tilde{t})$ in $p^e(\tilde{\theta}_i|s_i, \tilde{s}_i, t, \mathbb{E}(\tilde{t}))$:

$$p^e(\tilde{\theta}_i|s_i, \tilde{s}_i) \geq p^e(\tilde{\theta}_{-i}|s_i, \tilde{s}_{-i})$$

$$\begin{aligned}\Leftrightarrow & p(X_i|s_i, \tilde{s}_i)p(\tilde{\theta}_i|X_i, \tilde{s}_i) + p(X_{-i}|s_i, \tilde{s}_i)p(\tilde{\theta}_i|X_{-i}, \tilde{s}_i) \geq \\ & p(X_i|s_i, \tilde{s}_{-i})p(\tilde{\theta}_{-i}|X_i, \tilde{s}_{-i}) + p(X_{-i}|s_i, \tilde{s}_{-i})p(\tilde{\theta}_{-i}|X_{-i}, \tilde{s}_{-i})\end{aligned}$$

$$\begin{aligned}\Leftrightarrow & \left[(2\tau(\nu(\tilde{s}_i)) - 1)p(\theta_i|s_i, t) + 1 - \tau(\nu(\tilde{s}_i))\right] p(\tilde{\theta}_i|X_i, \tilde{s}_i) + \\ & \left[(1 - 2\tau(\nu(\tilde{s}_i)))p(\theta_i|s_i, t) + \tau(\nu(\tilde{s}_i))\right] p(\tilde{\theta}_i|X_{-i}, \tilde{s}_i) \geq \\ & \left[(2\tau(\nu(\tilde{s}_{-i})) - 1)p(\theta_i|s_i, t) + 1 - \tau(\nu(\tilde{s}_{-i}))\right] p(\tilde{\theta}_{-i}|X_i, \tilde{s}_{-i}) + \\ & \left[(1 - 2\tau(\nu(\tilde{s}_{-i})))p(\theta_i|s_i, t) + \tau(\nu(\tilde{s}_{-i}))\right] p(\tilde{\theta}_{-i}|X_{-i}, \tilde{s}_{-i})\end{aligned}$$

$$\begin{aligned}\Leftrightarrow & p(\theta_i|s_i, t)\Big\{\left[2\tau(\nu(\tilde{s}_i)) - 1\right](p(\tilde{\theta}_i|X_i, \tilde{s}_i) - p(\tilde{\theta}_i|X_{-i}, \tilde{s}_i)) + \\ & \left[2\tau(\nu(\tilde{s}_{-i})) - 1\right](p(\tilde{\theta}_{-i}|X_{-i}, \tilde{s}_{-i}) - p(\tilde{\theta}_{-i}|X_i, \tilde{s}_{-i}))\Big\} \geq \\ & (1 - \tau(\nu(\tilde{s}_{-i})))p(\tilde{\theta}_{-i}|X_i, \tilde{s}_{-i}) + \tau(\nu(\tilde{s}_{-i}))p(\tilde{\theta}_{-i}|X_{-i}, \tilde{s}_{-i}) - \\ & (1 - \tau(\nu(\tilde{s}_i)))p(\tilde{\theta}_i|X_i, \tilde{s}_i) - \tau(\nu(\tilde{s}_i))p(\tilde{\theta}_i|X_{-i}, \tilde{s}_i)\end{aligned}$$

Noticing that the right-hand side does not depend on the true ability of the reviewer, and that the expression between brackets in the left-hand side is positive (and does not depend on the true ability), we get that the inequality is more slack when $p(\theta_i|s_i, t)$ increases. Therefore, if the incentive compatibility is satisfied for a true ability $t^a \leq t^b$, it is also satisfied for t^b.

B Identification and Estimation Strategy of the Model with Potential Conflicts of Interest

As highlighted in DellaVigna & Kennedy (2012), potential conflicts of interest could arise in movie reviews if the reviewer's newspaper and the movie's production company belong to the same media outlet. An example would be a movie released by the 20^{th} Century Fox reviewed in the New York Post, both members of the media outlet Newscorp. We detail all possible conflicts of interest in the data section.

We include these conflict of interest in our structural estimation by modeling the behavior of the movie reviewer facing a movie from her own media outlet. We denote b^c her bias due to the conflict of interest and say that $b^c = p(r_1|\hat{r_0})$ in which $\hat{r_0}$ is the recommendation the reviewer would have sent in the absence of any conflict of interest. This is therefore equivalent to a situation in which the movie reviewer plays the game as described in the theoretical model but switches her recommendation with probability b^c if she were to send a negative recommendation concerning a movie from her media outlet.

The identification and estimation strategies are similar to those for the main model except for the reduced-form conditional grading probabilities, which change in the following way in case of a conflict of interest:

Table 8: Probabilities of giving a good review conditional on the true state of the world and on the prior in the case of a conflict of interest.

Prior	$\gamma_{i,1}^c$	$\gamma_{i,0}^c$
$\mu_j < \underline{\mu_i}$	$(1-t)b^c + t(1-b^- + b^- b^c)$	$tb^c + (1-t)(1-b^- + b^- b^c)$
$\mu_j \in \left[\underline{\mu_i}, \overline{\mu_i}\right]$	$t + (1-t)b^c$	$(1-t) + tb^c$
$\mu_j > \overline{\mu_i}$	$t + (1-t)(b^+ + (1-b^+)b^c)$	$(1-t) + t(b^+ + (1-b^+)b^c)$

We estimate this model by maximizing the following maximum likelihood:

$$\max_{\left\{t_i, b_i^-, b_i^+,\right\}_{i=1}^n, \beta, \delta, b^c} \sum_j \log \Bigg[\mu(\omega_j;\beta) \prod_{i=1}^n \Big\{ \left({\gamma_{i,1}^c}^{r_i} (1-\gamma_{i,1}^c)^{1-r_i} \right)^{C_{ij}} \left({\gamma_{i,1}}^{r_i} (1-\gamma_{i,1})^{1-r_i} \right)^{1-C_{ij}} \Big\}$$
$$+ (1-\mu(\omega_j;\beta)) \prod_{i=1}^n \Big\{ \left({\gamma_{i,1}^c}^{r_i} (1-\gamma_{i,1}^c)^{1-r_i} \right)^{C_{ij}} \left({\gamma_{i,1}}^{r_i} (1-\gamma_{i,1})^{1-r_i} \right)^{1-C_{ij}} \Big\} \Bigg]$$

$$\text{s.t. } t_i \in \left[\frac{1}{2}, 1\right], \quad b_i^+, b_i^-, b^c \in [0,1], \delta \geq 0 \qquad \text{(B.1)}$$

In this likelihood, C_{ij} is a dummy variable taking value 1 if there is a potential conflict of interest between movie reviewer i for movie j. $\gamma_{i,\theta}^c$ and $\gamma_{i,\theta}$ in this expression are taken respectively from Tables 8 and 1, their value depending on the prior. For instance, if $\mu(\omega_j;\beta) < \underline{\phi}(\mathbb{E}(t_i|\omega_i;\delta))$, the expressions for $\gamma_{i,\theta}^c$ and $\gamma_{i,\theta}$ are taken in the first line of the tables. Notice also that we do not estimate a bias due to conflict of interest which is reviewer-specific (although we could in theory), because of a lack of observations.

C Description of the Data

Table 9: Variables used in our estimations, their source, and description.

Variable	**Description**	**Source**
Title	Title of the movie	
Director	Name of the movie's director	
Year	Year of release of the movie in the USA	*www.imdb.com*

Variable	Description	Source
r_{ij}	The recommendation or review given by movie reviewer i to movie j, either 1 (fresh) or 0 (rotten)	*www.rottentomatoes.com*
Reviewer	Name of the author of the review	*www.rottentomatoes.com*
Newspaper	Name of the medium in which the review is published (Note: it is not necessarily a newspaper)	*www.rottentomatoes.com*
Review's Date	The date when the review was published	*www.rottentomatoes.com*
Number of previously directed films	Number of previous films by the director. In the case of several directors for one movie, the sum of the number of previous films.	*www.imdb.com*
Number of previous citations	Number of previous citations by the director: we count as a citation the fact that a movie is referenced or featured in another movie.	*www.imdb.com*
Budget	The estimated production budget of a movie in 2013's US dollars. We used historical data on exchange rates for non-US movies and the Consumer Price Index (CPI-U) data provided by the U.S. Department of Labor Bureau of Labor Statistic to take into account inflation.	*www.imdb.com*

Variable	Description	Source
Box-Office US	Gross Profit in the USA, expressed in 2013's US dollars	*www.imdb.com*
USA	Dummy equal to 1 if the movie is produced in the USA only.	*www.imdb.com*
Coproduction USA	Dummy equal to 1 if the movie is produced in the USA and at least another country.	*www.imdb.com*
Genre	A proxy for the type of the movie (e.g. action, thriller, documentary)	*www.imdb.com*
G, PG, PG-13, R, NC-17	MPAA rating of the movie	*www.imdb.com*
Remake	Dummy equal to 1 if the movie is a remake.	*www.imdb.com*
Sequel	Dummy equal to 1 if the movie is a sequel.	*www.imdb.com*
Google Search Index	Index taking a value 100 the month Peter Travers (our benchmark reviewer in reputation) got the most searches of his name on Google. All other values are computed related to this reference point. Note that these values are updated each day. We collected our data on reputation on May 6th, 2013.	*www.google.com/trends*
Production Company	Main production company of the movie (in general the distributor)	*www.metacritic.com*, *www.rottentomatoes.com*

Variable	Description	Source
Conflict of Interest	Dummy equal to 1 if Newspaper and Production Company belong to the same media outlet	DellaVigna & Kennedy (2011), *www.wikipedia.com*
Format	Format of the medium through which the review is published (e.g. newspaper, magazine, tabloid)	*www.wikipedia.com*
Content	Content of the medium through which the review is published (e.g. news, culture, cinema)	*www.wikipedia.com*
Target	Commercial target of the medium through which the review is published, i.e. whether it is designed for a general audience or professionals.	*www.wikipedia.com*

Table 10: Movie reviewers included in the empirical analysis along with their number of reviews in our sample, their average Google Search index through the sampling period, and the media through which they published.

Reviewer's Name	Total # Reviews	Avg. Google Search index	Media
Ann Hornaday	857	0.88	Washington Post
Ao Scott	1,235	6	At the Movies, New York Times
Carrie Rickey	1,020	0.045	Philadelphia Inquirer

Reviewer's Name	Total # Reviews	Avg. Google Search index	Media
Claudia Puig	1,469	0.67	USA Today
Colin Covert	1,435	0.1	Chicago Tribune, Minneapolis Star Tribune
David Edelstein	1,029	11.15	NPR, New York Magazine, Slate
Desson Thomson	1,266	0	Washington Post
Elizabeth Weitzman	997	0.09	New York Daily News
James Berardinelli	2,858	26.36	Reelviews
Joe Baltake	904	0	Passionate Moviegoer, Sacramento Bee
Kenneth Turan	1,124	3.89	Los Angeles Times, Newsday
Kirk Honeycutt	1,163	0.009	Hollywood Reporter
Kyle Smith	1,010	40.41	New york Post
Liam Lacey	998	0.2	Globe and Mail
Lisa Schwarzbaum	1,633	0.5	Entertainment Weekly
Lou Lumenick	1,723	0.63	New York Post
Michael Phillips	1,157	72.62	At the Movies, Chicago Tribune
Mick Lasalle	1,579	2.62	Houston Chronicle, San Francisco Chronicle
Moira Macdonald	1,415	0.18	Seattle Times
Owen Gleiberman	1,971	0.79	CNN.com, Entertainment Weekly
Peter Howell	1,260	9.75	Toronto Star
Peter Travers	1,879	23.24	Rolling Stone

Reviewer's Name	Total # Reviews	Avg. Google Search index	Media
Rex Reed	940	18.125	New York Observer
Richard Roeper	1,736	35.31	Chicago Sun-Times, Ebert & Roeper, Richard Roeper.com
Robert Denerstein	946	0	Denver Rocky Mountain
Roger Ebert	3,277	1100	Chicago Sun-Times, Denver Post, Detroit News, Ebert & Roeper
Stephanie Zacharek	944	1.21	CNN.com, Film.com, Los Angeles Times, NPR, Salon.com, Village Voice
Stephen Holden	1,050	5.17	New York Times
Stephen Whitty	1,403	0	Newark Star-Ledger
Steven Rea	1,232	1.18	Philadelphia Inquirer
Terry Lawson	1,193	3.34	Detroit Free Press, Miami Herald
Todd Mccarthy	903	3.125	Hollywood Reporter, Variety, indieWire
Ty Burr	1,313	1.375	Boston Globe, Dallas Morning News, Entertainment Weekly
Unknown Reviewer	2,142	0	Time Out (40%)
Wesley Morris	1,217	3.49	Boston Globe

Table 11: Potential Conflicts of Interest

Media Outlet	Production Companies	Media	Years of Interactions	Concerned Reviewers	# Observations
News Corp.	20th Century Fox, Fox Searchlight Pictures	New York Post	1993-2013	Kyle Smith	56
				Lou Lumenick	74
Time Warner	Warner Bros Pictures, Picturehouse, HBO	Entertainment Weekly	1990-2013	Lisa Schwarzbaum	88
				Owen Gleiberman	102
				Unknown Reviewer	3
		TIME Magazine	1990-2013	Unknown Reviewer	1
	New Line	Entertainment Weekly	1996-2010	Lisa Schwarzbaum	18
				Owen Gleiberman	31
		TIME Magazine	1996-2010	Unknown Reviewer	1
	Fine Line Features	Entertainment Weekly	1996-2013	Lisa Schwarzbaum	4
				Owen Gleiberman	3
Disney Media Group	Walt Disney Pictures, Buena Vista	At the Movies	2007-2010	Ao Scott	5
				Michael Phillips	1
		Ebert & Roeper	2007-2013	Richard Roeper	40
	Miramax	At the Movies	2007-2010	Ao Scott	2
				Michael Phillips	1
		Ebert & Roeper	2007-2010	Richard Roeper	22

D Additional Tables

Table 12: ML estimates for reviewer-specific parameters — Specification (II) includes the budget and conflicts of interest

	(I)						(II)					
	t		b^-		b^+		t		b^-		b^+	
Reviewer's Name	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Ann Hornaday	0.781	(0.021)	0.027	(0.045)	0.094	(0.047)	0.785	(0.029)	0.082	(0.059)	0.143	(0.063)
Ao Scott	0.758	(0.017)	0.001	(0.034)	0.175	(0.051)	0.741	(0.024)	0.032	(0.051)	0.089	(0.056)
Carrie Rickey	0.784	(0.015)	0.000	(0.000)	0.367	(0.066)	0.767	(0.027)	0.000	(0.015)	0.349	(0.062)
Claudia Puig	0.822	(0.006)	0.073	(0.038)	0.195	(0.041)	0.833	(0.018)	0.124	(0.047)	0.140	(0.041)
Colin Covert	0.765	(0.009)	0.002	(0.000)	0.184	(0.043)	0.773	(0.018)	0.002	(0.001)	0.235	(0.050)
David Edelstein	0.769	(0.014)	0.001	(0.030)	0.250	(0.055)	0.773	(0.025)	0.062	(0.044)	0.223	(0.057)
Desson Thomson	0.821	(0.023)	0.188	(0.072)	0.001	(0.028)	0.819	(0.029)	0.155	(0.083)	0.001	(0.031)
Elizabeth Weitzman	0.804	(0.015)	0.024	(0.051)	0.065	(0.041)	0.815	(0.024)	0.081	(0.061)	0.075	(0.047)
James Berardinelli	0.768	(0.011)	0.155	(0.040)	0.110	(0.030)	0.772	(0.019)	0.172	(0.043)	0.077	(0.046)
Joe Baltake	0.657	(0.043)	0.000	(0.000)	0.432	(0.090)	0.663	(0.052)	0.000	(0.004)	0.510	(0.101)
Kenneth Turan	0.789	(0.017)	0.007	(0.035)	0.233	(0.060)	0.758	(0.033)	0.030	(0.046)	0.217	(0.086)
Kirk Honeycutt	0.734	(0.014)	0.029	(0.044)	0.095	(0.051)	0.734	(0.026)	0.039	(0.044)	0.045	(0.048)
Kyle Smith	0.619	(0.015)	0.345	(0.047)	0.000	(0.000)	0.654	(0.026)	0.360	(0.065)	0.000	(0.000)
Liam Lacey	0.790	(0.026)	0.406	(0.055)	0.004	(0.000)	0.794	(0.029)	0.439	(0.065)	0.004	(0.006)
Lisa Schwarzbaum	0.823	(0.009)	0.000	(0.030)	0.164	(0.056)	0.823	(0.024)	0.016	(0.042)	0.132	(0.048)
Lou Lumenick	0.821	(0.006)	0.202	(0.044)	0.046	(0.025)	0.823	(0.021)	0.178	(0.060)	0.003	(0.042)
Michael Phillips	0.799	(0.006)	0.092	(0.035)	0.091	(0.037)	0.789	(0.023)	0.111	(0.053)	0.075	(0.050)
Mick Lasalle	0.665	(0.011)	0.018	(0.039)	0.147	(0.050)	0.661	(0.024)	0.003	(0.035)	0.119	(0.051)
Moira Macdonald	0.838	(0.000)	0.003	(0.009)	0.241	(0.052)	0.836	(0.021)	0.003	(0.036)	0.213	(0.053)
Owen Gleiberman	0.768	(0.011)	0.000	(0.000)	0.263	(0.038)	0.770	(0.025)	0.000	(0.008)	0.267	(0.049)
Peter Howell	0.807	(0.014)	0.008	(0.035)	0.234	(0.049)	0.811	(0.022)	0.045	(0.043)	0.273	(0.054)
Peter Travers	0.877	(0.006)	0.052	(0.028)	0.357	(0.046)	0.857	(0.019)	0.059	(0.038)	0.265	(0.052)
Rex Reed	0.670	(0.016)	0.129	(0.068)	0.001	(0.027)	0.680	(0.029)	0.224	(0.075)	0.001	(0.018)
Richard Roeper	0.791	(0.009)	0.008	(0.006)	0.355	(0.044)	0.787	(0.020)	0.004	(0.015)	0.266	(0.053)
Robert Denerstein	0.898	(0.009)	0.107	(0.035)	0.001	(0.034)	0.898	(0.021)	0.146	(0.062)	0.013	(0.042)
Roger Ebert	0.782	(0.000)	0.000	(0.000)	0.365	(0.035)	0.775	(0.015)	0.000	(0.007)	0.353	(0.047)
Stephanie Zacharek	0.700	(0.020)	0.080	(0.048)	0.004	(0.011)	0.688	(0.028)	0.115	(0.062)	0.004	(0.013)
Stephen Holden	0.790	(0.020)	0.057	(0.051)	0.102	(0.040)	0.800	(0.032)	0.076	(0.085)	0.014	(0.052)
Stephen Whitty	0.722	(0.006)	0.255	(0.053)	0.004	(0.000)	0.726	(0.022)	0.276	(0.052)	0.004	(0.002)
Steven Rea	0.771	(0.011)	0.023	(0.041)	0.227	(0.056)	0.763	(0.025)	0.164	(0.065)	0.244	(0.055)
Terry Lawson	0.799	(0.020)	0.104	(0.050)	0.186	(0.053)	0.795	(0.029)	0.151	(0.067)	0.147	(0.060)
Todd Mccarthy	0.782	(0.012)	0.039	(0.039)	0.109	(0.047)	0.790	(0.028)	0.019	(0.054)	0.095	(0.061)
Ty Burr	0.791	(0.009)	0.047	(0.034)	0.000	(0.021)	0.772	(0.022)	0.051	(0.045)	0.000	(0.026)
Unknown Reviewer	0.749	(0.016)	0.080	(0.056)	0.055	(0.051)	0.757	(0.039)	0.217	(0.102)	0.109	(0.073)
Wesley Morris	0.738	(0.006)	0.318	(0.055)	0.005	(0.000)	0.742	(0.024)	0.330	(0.054)	0.001	(0.000)
# Observations:	30440						22674					

Notes: Bootstrap Standard Errors are computed on 100 iterations for (I) and 500 for (II)

Table 13: ML Estimates for Movie Specific Parameters on a Subset of the Data Set — Specification (II) is defined as in section 5; Specification (III) excludes strategic biases; Specification (IV) is a naive logit estimation of the prior.

	(II)		(III)		(IV)	
	Coeff.	SE	Coeff.	SE	Coeff.	SE
Movie Specific, β:						
Constant	1.124	(0.010)	3.076	(0.842)	2.208	(0.213)
Origin: USA	-1.440	(0.037)	-0.834	(0.226)	-0.433	(0.056)
Origin: co-production USA	-1.556	(0.102)	-0.823	(0.244)	-0.428	(0.059)
Remake	-0.440	(0.232)	-0.594	(0.270)	-0.256	(0.062)
Sequel	-0.434	(0.140)	-0.263	(0.210)	-0.076	(0.049)
Director's number of previous movies	0.016	(0.000)	0.029	(0.010)	0.017	(0.002)
G rating	0.953	(3.439)	1.704	(1.224)	0.972	(0.136)
PG rating	0.470	(0.201)	0.151	(0.206)	0.151	(0.049)
R rating	0.530	(0.060)	0.361	(0.129)	0.219	(0.035)
NC-17 rating	0.345	(6.490)	0.077	(1.976)	0.165	(0.222)
log(budget)	0.000	(0.000)	-0.157	(0.050)	-0.111	(0.013)
Reputation, δ:						
Google Search Index	3.5×10^{-4}	(0.000)				
Conflict of interest, b^c:						
Average Bias	0.043	(0.040)				
Likelihood	-10422		-10620		-12144	
# Observations:	17882		17882		17882	

Notes: Standard Errors are bootstrapped on 189 iterations for (II) and 200 for (III).

Table 14: ML Estimates for Reviewer-Specific Parameters on a Subset of the Data Set — Specification (II) is defined as in section 5; Specification (III) excludes strategic biases.

	(II)						(III)	
	t		b^-		b^+		t	
Reviewer's Name	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Ann Hornaday	0.782	(0.029)	0.125	(0.067)	0.205	(0.066)	0.768	(0.023)
Ao Scott	0.756	(0.028)	0.020	(0.048)	0.097	(0.060)	0.749	(0.022)
Carrie Rickey	0.760	(0.029)	0.000	(0.014)	0.296	(0.083)	0.754	(0.023)
Claudia Puig	0.822	(0.020)	0.108	(0.051)	0.128	(0.051)	0.805	(0.014)
Colin Covert	0.768	(0.020)	0.000	(0.001)	0.224	(0.051)	0.746	(0.017)
David Edelstein	0.775	(0.025)	0.039	(0.046)	0.207	(0.060)	0.770	(0.021)
Desson Thomson	0.806	(0.032)	0.154	(0.093)	0.000	(0.021)	0.804	(0.027)
Elizabeth Weitzman	0.792	(0.030)	0.084	(0.074)	0.093	(0.064)	0.778	(0.021)
James Berardinelli	0.784	(0.017)	0.164	(0.052)	0.098	(0.048)	0.762	(0.015)
Joe Baltake	0.644	(0.062)	0.000	(0.011)	0.515	(0.106)	0.652	(0.045)
Kenneth Turan	0.769	(0.032)	0.027	(0.056)	0.215	(0.090)	0.768	(0.026)
Kirk Honeycutt	0.727	(0.026)	0.000	(0.048)	0.071	(0.061)	0.717	(0.023)
Kyle Smith	0.646	(0.028)	0.361	(0.071)	0.000	(0.000)	0.630	(0.025)
Liam Lacey	0.788	(0.029)	0.461	(0.076)	0.000	(0.011)	0.744	(0.025)
Lisa Schwarzbaum	0.822	(0.022)	0.050	(0.053)	0.111	(0.060)	0.802	(0.019)
Lou Lumenick	0.825	(0.023)	0.170	(0.071)	0.000	(0.047)	0.799	(0.018)
Michael Phillips	0.796	(0.021)	0.122	(0.056)	0.070	(0.053)	0.783	(0.016)
Mick Lasalle	0.655	(0.022)	0.000	(0.038)	0.123	(0.068)	0.657	(0.022)
Moira Macdonald	0.821	(0.024)	0.000	(0.025)	0.207	(0.063)	0.807	(0.018)
Owen Gleiberman	0.766	(0.025)	0.000	(0.024)	0.290	(0.054)	0.738	(0.024)
Peter Howell	0.807	(0.025)	0.042	(0.047)	0.263	(0.060)	0.769	(0.020)
Peter Travers	0.849	(0.022)	0.078	(0.051)	0.262	(0.059)	0.825	(0.016)
Rex Reed	0.670	(0.030)	0.280	(0.081)	0.000	(0.001)	0.620	(0.028)
Richard Roeper	0.785	(0.022)	0.000	(0.008)	0.257	(0.059)	0.749	(0.019)
Robert Denerstein	0.898	(0.022)	0.201	(0.069)	0.019	(0.043)	0.848	(0.024)
Roger Ebert	0.778	(0.020)	0.000	(0.010)	0.337	(0.054)	0.761	(0.015)
Stephanie Zacharek	0.683	(0.027)	0.078	(0.071)	0.000	(0.002)	0.656	(0.027)
Stephen Holden	0.783	(0.036)	0.088	(0.090)	0.026	(0.056)	0.793	(0.028)
Stephen Whitty	0.698	(0.020)	0.297	(0.070)	0.000	(0.000)	0.670	(0.019)
Steven Rea	0.765	(0.029)	0.118	(0.070)	0.226	(0.055)	0.740	(0.022)
Terry Lawson	0.761	(0.032)	0.146	(0.080)	0.200	(0.069)	0.737	(0.023)
Todd Mccarthy	0.791	(0.030)	0.000	(0.051)	0.106	(0.071)	0.780	(0.025)
Ty Burr	0.782	(0.021)	0.051	(0.061)	0.000	(0.032)	0.769	(0.019)
Unknown critic	0.759	(0.041)	0.237	(0.108)	0.101	(0.083)	0.750	(0.033)
Wesley Morris	0.744	(0.028)	0.335	(0.065)	0.000	(0.000)	0.717	(0.021)
# Observations:			17882				17882	

Notes: Bootstrap Standard Errors are computed on 189 iterations for (II) and 200 for (III).

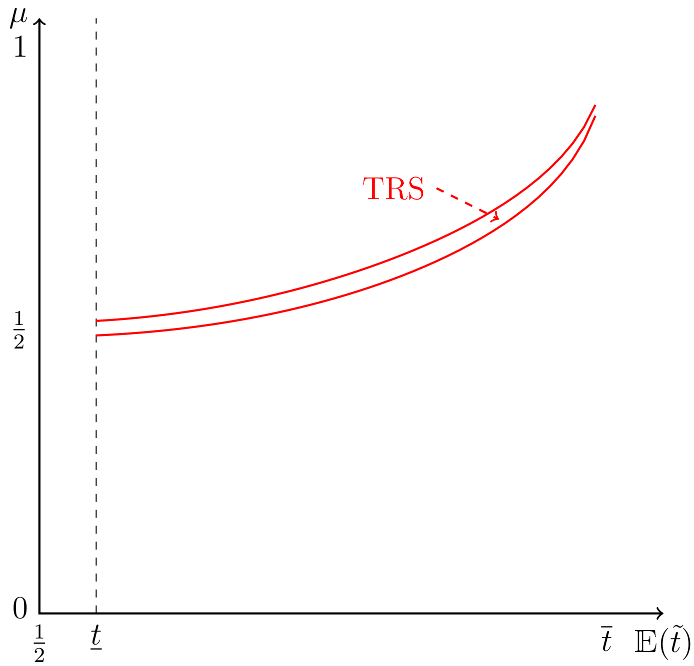

Figure 5: Truthful revelation set for $\underline{t} = 0.55$ and $\tau(\nu) = \frac{1+\nu}{2}$.